Health Evidence Review Commission's

Evidence-based Guidelines Subcommittee

October 4, 2012

Wilsonville Training Center, Room 112
Clackamas Community College
29353 SW Town Center Loop E
Wilsonville, Oregon 97070

AGENDA
EVIDENCE-BASED GUIDELINES SUBCOMMITTEE (EbGS)
October 4, 2012
2:00pm - 5:00pm
Wilsonville Training Center *Room 112*

(All agenda items are subject to change and times listed are approximate)

#	Time	Item	Presenter
1	2:00 PM	Call to Order	Steve Marks
2	2:05 PM	Review of August minutes	Steve Marks
3	2:10 PM	Coverage Guidance Evidence Algorithm	Cat Livingston
4	2:45 PM	Review public comment and finalize CGs: 1. Management of chronic otitis media in children 2. Percutaneous interventions for low back pain	Alison Little Cat Livingston
5	3:20 PM	New Draft Coverage Guidances 1. Advanced imaging for cardiac disease A. Coronary artery calcium scoring B. Coronary computed tomography angiography 2. Management of recurrent acute otitis media in children 3. Cervical cancer screening	Cat Livingston
6	4:35 PM	Confirmation of next meeting December 6, 2012	Steve Marks
7	4:40 PM	Other business	
8	4:45 PM	Public Comment	
9	4:55 PM	Next topics	Cat Livingston
10	5:00 PM	Adjournment	Steve Marks

MINUTES

Evidence-based Guidelines Subcommittee

Meridian Park Room
Community Health Education Center, Room 117 B&C
19300 SW 65th Avenue, Tualatin, OR 97062
August 2, 2012
2:00pm - 5:00pm

Members Present: Wiley Chan, MD, Chair; Vern Saboe, DC; Irene Croswell, RPh; Leda Garside, RN; Bob Joondeph, JD; Eric Stecker, MD.

Members Absent: Som Saha, MD, MPH; Beth Westbrook, PsyD; Steve Marks, MD, Vice-Chair.

Staff Present: Darren Coffman, MS; Cat Livingston, MD, MPH.

Also Attending: Alison Little, MD and Shannon Vandegriff, Center for Evidence-based Policy (CEbP); Chandra Basham, MD and Denise Taray (DMAP); Andrea Herska, MD.

Roll Call/Minutes Approval/Staff Report

The meeting was called to order at 2:04 pm and roll was called. Dr. Chan introduced the agenda. Minutes from the June 7, 2012 EbGS meeting were reviewed. It was clarified that Bob Joondeph was present, and the modified minutes were approved.

ACTION: HERC staff will post the approved minutes on the website as soon as possible.

Staff report
Dr. Livingston reported on the potential topic of Neck Pain that was proposed at the last EbGS meeting. There was no single evidence based systematic review, but there were a series of Cochrane reviews that addressed individual interventions. Given the lack of clear evidence base, it would be difficult to turn this into a coverage guidance, and may be more appropriate for a guideline topic. Vern Saboe asked about the process of choosing guideline topics and this was clarified.

Topic: Review public comment and finalize CGs.

- **Femoroacetabular Impingment Syndrome Surgery**

Discussion: Little presented the public comment disposition chart and there was discussion about the comments and recommendations.

Public comment was received. Dr. Andrea Herska, an orthopedic surgeon who practices at OHSU testified that coverage of FAI surgery was recommended. She disclosed that

she has a potential conflict of interest, as she instructs for Smith and Nephew on FAI exams.

Dr. Herska provided background on FAI. Twenty percent of anyone walking down the street, have hips shaped with FAI. A large percentage will go on to require hip replacement surgery in 50s. She acknowledged that we don't know if surgery will alter the natural history of the disease. Most insurance companies now don't require physical therapy when there is a labral tear. She states it unethical to perform an RCT because these are symptomatic individuals.

There was a discussion of the lack of controlled trials. Dr. Stecker raised the issue of haphazard input from associations. Bob Joondeph discussed the role of the EbGS, as to reviewing evidence or obtaining input. Stecker argues for creating coverage guidance acknowledging standard of care. In this case, if the evidence is insufficient, but it is in widespread practice, it may be appropriate for coverage to continue.

Dr. Chan had also spoken with the Kaiser Permanente head of orthopedics at the national level who also concurred that the prevalent opinion of orthopedists is to treat FAI with surgery.

A discussion was made of what the options were. To make a coverage recommendation with insufficient evidence, to make a noncoverage recommendation that goes against common practice, or to table a recommendation based on the apparent conflict with no other good treatment options. The decision was made to table a decision.

Actions:

1) Recommend HERC table development of a coverage guidance for Hip Surgery for FAI
2) Staff to contact DMAP and gain insight into PA process, and potentially address this issue with the Value-based Benefits Subcommittee
3) Staff to follow up with Dr. Herzka to explain how this issue may be reviewed at the Value-based Benefits Subcommittee

Motion to approve recommendation to table coverage guidance. Motion carries: 6-0.

- **Topic: Other business - General principles and decision making in EbGS**

Discussion: Bob Joondeph raised that there should be a decision tree developed to ensure that a consistent approach to developing coverage guidance exists. The general principles and questions that the group discussed included:

- If only have yellow light categories – consider tabling coverage guidance
- If a conflict between two or more trusted sources – how to weigh them differently and how does common practice and alternative availability impact the decision. This was determined to require case-by-case review
- If insufficient evidence exists, and there is no other viable alternative, but a treatment is widely practiced, consider making no recommendation and having this reviewed by VbBS instead.
- If common practices differs from the evidence, the evidence needs to be strong enough to make a recommendation that conficts with common practice

- Standard of care alone should not guide a coverage decision
- If not enough evidence and no professional consensus, then red light decision
- If not enough evidence and professional consensus, then table it
- Insufficient evidence, no alternative, consider coverage
- Insufficient evidence, good alternative, don't cover

There was a discussion about standard of care versus professional consensus. There was clarification that the legal definition may or may not be applicable in this circumstance. There was a concern raised that professional societies may be biased and not all of them consistently use rigorous evidence-based decision making in their guideline development.

Actions:

1) CEbP and HERC Staff to develop a draft decision algorithm to capture the principles shared and bring this to the October meeting.

➢ **Topic: Advanced Imaging for Low Back Pain**

Discussion: Little presented the public comment disposition chart. There was discussion with a decision to adopt the dispositions as recommended, however, there were some outstanding concerns about Table B. Specifically, the lack of mention of immunosuppression was concerning and confusion about why those specific features of cancer were chosen. It was clarified that this was initially directly from the evidence-based guideline, and already modified somewhat. The decision was to keep a list of potential changes for Table B and potentially revise it in the future. Since electrophysiologic testing was considered useful in diagnosis, it was decided to strike this from the coverage box altogether (remain silent).

Actions:

1) Table B – in future consider making changes to
    a. Cancer feature language
    b. Add immunosuppression to row 2
2) Send Table B recommendations to ACP
3) Approve the following final draft coverage guidance to forward to HERC:
    In patients with non-specific low back pain and an absence of "red flag" conditions:
    1) imaging (e.g. x-ray, CT, or MRI) should not be covered, and
    2) thermography and surface electromyography should not be covered.

    If patients have persistent (> 1 month) low back pain and signs or symptoms of radiculopathy or spinal stenosis, appropriate imaging should be covered (MRI preferred, CT second choice) **only if** the patient is a potential candidate for surgery or epidural steroid injection (for suspected radiculopathy).

    If patients have severe or progressive neurologic deficits, or clinicians suspect a serious underlying condition (e.g. cancer or infection) prompt workup with MRI (first choice) or CT should be covered.

Motion to approve draft coverage guidance as amended for HERC consideration. Motion carries: 6-0.

- **Topic: Non-Pharmacologic Interventions For Treatment Resistant Depression**

  Discussion: Little presented the public comment disposition chart. There was some discussion about the role of electroconvulsive therapy (ECT). Further clarification was made that this does not address the initial treatment of depression, only treatment resistant depression. Livingston suggested a clarifying sentence added below the coverage guidance box to this effect, which was adopted. There was a concern that ECT is controversial, which resulted in one abstention.

  Actions:

  1) Add clarifying sentence as footnote to coverage guidance box: "Coverage guidance for initial therapy, which may include pharmacologic treatment and/or psychotherapy, is outside the scope of this document."
  2) Approve the following final draft coverage guidance to forward to HERC:
     In patients with an episode of major depressive disorder who have failed an initial trial of antidepressants, psychotherapy should be covered.

     In patients with an episode of major depressive disorder who have failed at least two or more pharmacologic treatments the following treatments should be covered:
     1) Repetitive transcranial magnetic stimulation
     2) Electroconvulsive therapy

     Vagus nerve stimulation should not be covered.

Motion to approve draft coverage guidance as amended for HERC consideration. Motion carries: 5-0 (Abstention: Joondeph).

- **Topic: Neuroimaging in dementia**

  Discussion: Little shared that there was no public comment received. There was no discussion.

  Actions:

  1) Approve the following final draft coverage guidance to forward to HERC:
     Screening of asymptomatic patients for dementia with neuroimaging should not be covered.

     Structural neuroimaging should be covered to rule out reversible causes of dementia.*

     In patients with mild cognitive impairment, imaging should not be used to predict progression of the risk of developing dementia.

     Functional neuroimaging (PET, SPECT or fMRI) should not be covered for screening, diagnosis, or monitoring of dementia.

Motion to approve draft coverage guidance as written for HERC consideration. Motion carries: 6-0.

- **Topic: Neuroimaging in headache**
  Discussion: Little shared that there was no public comment received. There was no discussion.

  Actions:
  2) Approve the following final draft coverage guidance to forward to HERC:
     In patients who present with a variation of their usual headache (e.g. more severe, longer in duration, or not responding to drugs), CT or MRI should not be covered.

     Neuroimaging (CT or MRI) should not be covered for those with a clear history of migraine, unless a red flag* is present.

Motion to approve draft coverage guidance as written for HERC consideration. Motion carries: 6-0.

- **Topic: New Draft Coverage Guidances: Pressure equalization (tympanostomy tubes) in children**

  Discussion:
  Livingston presented the proposed coverage guidance. There was a discussion about the title and the discrepancy between coverage guidance of acute otitis media and chronic otitis media. It was felt there should be greater consistency, either both address antibiotics or neither. Members decided to create two separate guidances, one on the Management of Chronic Otitis Media in Children, and the other on acute otitis media (AOM). For the AOM coverage guidance, if CEbP is able to identify an evidence based guidance on overall management that can be developed into a coverage guidance, that will be done. Otherwise, the AOM coverage guidance will just focus on tympanostomy tubes for AOM.

  There was an extensive discussion about the hearing decibel cutoff. Whether 20, 25 or 40 db should be used as a cutoff. NICE had used 25, and the MED report had identified studies suggesting that above this level there is a significant impact on effusion. The US guidelines had a 20-40 range that involves individualized treatment. There was a discussion about US common practice and where the evidence lies. The decision was made to adopt the 25dB cutoff rather than giving an option (of 20 or 40 in addition to 25dB).

  Action
  1) Rename this coverage guidance: Management of Chronic Otitis Media with Effusion
  2) Staff and CEbP to identify if adequate guidance exists to develop a coverage guidance on management of acute otitis media. If not, a separate guidance on pressure equalization tubes for acute otitis media in children with an expanded focused evidence summary will be brought back to the next meeting.
  3) Delete the third paragraph of the initial guidance
  4) Approve the following draft coverage guidance language for public comment:
     Antibiotic and other medication therapy (including antihistamines, decongestants, and nasal steroids) should not be covered for children with OME (without another appropriate diagnosis).

     There should be a 3 to 6 month watchful waiting period after diagnosis of otitis media

with effusion, and if documented hearing loss is ≥ 25dB in the better hearing ear, referral for tympanstomy surgery should be covered.

Formal audiometry should be covered for children with chronic OME present for 3 months or longer. Children with language delay, learning problems, or significant hearing loss should have hearing testing covered initially upon diagnosis. Children with chronic OME who are not at risk for language or developmental delay should be reexamined at 3- to 6-month intervals until the effusion is no longer present, significant hearing loss is identified, or structural abnormalities of the eardrum or middle ear are suspected.

Adenoidectomy should be covered in children over 3 years, with chronic OME who are having their second set of tubes. Adenoidectomy should not be covered at the time of the first PE tube insertion.

Patients with craniofacial anomalies, Down's syndrome, cleft palate, and patients with speech and language delay along with hearing loss should have coverage based on an individualized treatment plan.

Motion to approve draft coverage guidance as amended for public comment. Motion carries: 6-0.

- **Topic: New Draft Coverage Guidances: Percutaneous Interventions for Low Back Pain**

Discussion: Livingston presented the proposed coverage guidance. There was a discussion about what decision to make when insufficient evidence is available. Given that there is evidence of short-to-moderate benefit for at least one treatment for radicular low back pain, other types of treatments for which there are randomized controlled trials (RCTs) showing no benefit (e.g. 9 RCTs of radiofrequency denervation demonstrating no benefit), or for therapies for which there are no RCTs (coblation nuceloplasty), a decision of a non-coverage recommendation was made.

For non-radicular low back pain, the options were similarly debated. There is evidence of other modalities being useful in the treatment of non-radicular low back pain that have been addressed in a separate state guideline (Evaluation and Management of Low Back Pain). A non-coverage recommendation was again made for those therapies for which there is insufficient evidence of benefit.

Action

1) Approve the following draft coverage guidance language for public comment:

For radicular low back pain, Epidural steroid injections should be covered for patients with persistent radiculopathy due to herniated lumbar disc; it is recommended that shared decision-making regarding epidural steroid injection include a specific discussion about inconsistent evidence showing moderate short-term benefits, and lack of long-term benefits. If an epidural steroid injection does not offer benefit, repeated injections should not be covered.

Epidural steroid injections should NOT be covered for spinal stenosis.

For radicular low back pain, the following treatments should NOT be covered:

- coblation nuceleoplasty
- radiofrequency denervation

For nonradicular low back pain, the following treatments should NOT be covered:

- facet joint corticosteroid injection
- prolotherapy
- intradiscal corticosteroid injection
- local injections
- botulinum toxin injection
- epidural steroid injection
- intradiscal electrothermal therapy (IDET)
- therapeutic medial branch block
- radiofrequency denervation
- sacroiliac joint steroid injection
- coblation nucleoplasty
- percutaneous intradiscal radiofrequency thermocoagulation

Motion to approve draft coverage guidance as written for public comment. Motion carries: 6-0.

- **Next meeting:**

The next meeting was confirmed for October 4, 2012. Chan indicated he will not be present at the October meeting. Staff will check in with Steve Marks to see if he can Chair the October meeting.

- **Public Comment:**

No additional public comment was given.

- **Adjournment:**

The meeting was adjourned at 4:59 pm.

Review Public Comments

HEALTH EVIDENCE REVIEW COMMISSION (HERC)

DRAFT COVERAGE GUIDANCE: MANAGEMENT OF CHRONIC OTITIS MEDIA WITH EFFUSION IN CHILDREN

DATE: XX/XX/XXXX

HERC COVERAGE GUIDANCE

Antibiotic and other medication therapy (including antihistamines, decongestants, and nasal steroids) should not be covered for children with children with otitis media with effusion (OME) (without another appropriate diagnosis).

There should be a 3 to 6 month watchful waiting period after diagnosis of otitis media with effusion, and if documented persistent hearing loss is greater than or equal to 25dB in the better hearing ear, referral for tympanstomy surgery ~~should~~ may be covered, given short, but not long-term, improvement in hearing.

Formal audiometry should be covered for children with chronic OME present for 3 months or longer. Children with language delay, learning problems, or significant hearing loss should have hearing testing covered initially upon diagnosis. Children with chronic OME who are not at risk for language or developmental delay should be reexamined at 3- to 6-month intervals until the effusion is no longer present, significant hearing loss is identified, or structural abnormalities of the eardrum or middle ear are suspected.

~~Adenoidectomy should be covered in children over 3 years, with chronic OME who are having their second set of tubes.~~ Adenoidectomy should not be covered at the time of the first pressure equalization tube insertion.

Patients with craniofacial anomalies, Down's syndrome, cleft palate, and patients with speech and language delay along with hearing loss should have coverage based on an individualized treatment plan.

* Coverage guidance for acute otitis media will be addressed in a separate document.

RATIONALE FOR GUIDANCE DEVELOPMENT

The HERC selects topics for guideline development or technology assessment based on the following principles:

- Represents a significant burden of disease
- Represents important uncertainty with regard to efficacy or harms
- Represents important variation or controversy in clinical care
- Represents high costs, significant economic impact
- Topic is of high public interest

Oregon Health Authority

Coverage guidance development follows to translate the evidence review to a policy decision. Coverage guidance may be based on an evidence-based guideline developed by the Evidence-based Guideline Subcommittee or a health technology assessment developed by the Heath Technology Assessment Subcommittee. In addition, coverage guidance may utilize an existing evidence report produced by one of HERC's trusted sources, generally within the last three years.

EVIDENCE SOURCES

Effros, R., & Little, A. (2010). *Pressure equalization tubes in children.* Portland, OR: Center for Evidence-based Policy, Oregon Health & Science University.

Sources Cited in MED Report:

American Academy of Family Physicians, American Academy of Otolaryngology – Head and Neck Surgery, & American Academy of Pediatrics (AAFP/AAOHNS/AAP) Subcommittee on Otitis Media with Effusion. (2004). Clinical Practice Guideline: Otitis Media with Effusion. *Pediatrics, 113*(5), 1412-1429.

Griffin, G., Flynn, C.A., Bailey, R.E., & Schultz, J.K. (2006). Antihistamines and/or decongestants for otitis media with effusion (OME) in children. *Cochrane Database of Systematic Reviews*, *4*(CD003423), 1-44.

Kay, D.J., Nelson, M., & Rosenfeld, R.M. (2001). Meta-analysis of tympanostomy tube sequelae. *Otolaryngology-Head and Neck Surgery*, *124*, 374-380.

Leach, A.J., & Morris, P.S. (2006). Antibiotics for the prevention of acute and chronic suppurative otitis media in children. *Cochrane Database of Systematic Reviews*, *4*(CD004401), 1-70.

Lous, J., Burton, M.J., Felding, J., Ovesen, T., Rovers, M., & Williamson, I. (2005). Grommets (ventilation tubes) for hearing loss associated with otitis media with effusion in children. *Cochrane Data Base of Systematic Reviews*, *1* (CD001801), 1-58.

Mandel, E.M., & Casselbrant, M.L. (2006). Recent developments in the treatment of otitis media with effusion. *Drug, 66*(12), 1545-1576.

McDonald, S., Langton Hewer, C.D., & Nunez, D.A. (2008). Grommets (ventilation tubes) for recurrent acute otitis media in children. *Cochrane Database of Systematic Reviews*, *4*(CD 004741), 1-14.

National Collaborating Centre for Women's and Children's Health. (2008). Surgical management of otitis media with effusion in children. London: National Institute for Health and Clinical Excellence (NICE). Retrieved July 6, 2012, from www.nice.org.uk/nicemedia/pdf/CG60NICEguideline.pdf

DRAFT

Perera, R., Haynes, J., Glasziou, P.P., & Heneghan, C.J. (2006). Autoinflation for hearing loss associated with otitis media with effusion. *Cochrane Database of Systematic Reviews, 4*(CD006285), 1-28.

Rovers, M.M., Black, N., Browning, G.G., Maw, R., Zielhuis, G.A., & Haggard, M.P. (2005). Grommets in otitis media with effusion: an individual patient data meta-analysis. *Archives of Diseases of Childhood*, *90*(5), 480-485.

Simpson, S.A., Thomas, C.L., van der Linden, M., MacMillan, H., van der Wouden, J.C., & Butler, C.C. (2007). Identification of children in the first four years of life for early treatment for otitis media with effusion. *Cochrane Database of Systemic Reviews*, *1*(CD004163), 1-24.

Thomas, C.L., Simson, S., Butler, C., & van der Voort, J. (2006). Oral or topical nasal steroids for hearing loss associated with otitis media with effusion in children. *Cochrane Database of Systemic Reviews*, *3*(CD001935), 1-26.

The summary of evidence in this document is derived directly from these evidence sources, and portions are extracted verbatim.

SUMMARY OF EVIDENCE

Clinical Background

Otitis media is one of the most frequent infections in children and is a leading cause of both visits to the physician and use of antibiotics in this population. The direct costs of otitis media are estimated at $3 to 5 billion per year in the US. Recurrent infections or chronic fluid in the middle ear can cause hearing deficits, and there is concern that in a rapidly developing child, this could lead to language and other developmental problems.

Pressure equalization (PE) tubes are small plastic or metal tubes that are surgically inserted into the tympanic membrane to allow for drainage of the fluid from the middle ear with the goal of improved hearing. The hope is that if hearing is improved, then language and other developments can be optimized. One of the challenges of determining which children require PE tube placement is that not all middle ear disease is associated with hearing loss, and even the presence of a mild to moderate hearing loss from a middle ear effusion does not necessarily translate into later speech or language delays in children. Further, the high rates of spontaneous resolution of both acute otitis media and middle ear effusions, and the fact that most PE tubes only remain in the ear drum for 6-12 months, may lessen the potential benefit of PE tube insertion.

Evidence Review

There is evidence that PE tubes decrease the duration of otitis media with effusion (OME) over the first year. In addition, PE tubes provide short-term (three to six month)

DRAFT

improvements in hearing, but this advantage dissipates by 12 months. Overall, there do not seem to be consistent benefits in language and development as a result of PE tube placement for OME. The most common complication of PE tubes appears to be otorrhea, which can result in increased use of oral or topical antibiotics. Tympanosclerosis and retraction pockets of the tympanic membrane are also complications of PE tubes, but their clinical significance remains uncertain. Limited evidence suggests that children with PE tubes sustain higher costs in follow-up, in addition to the costs of the procedure itself, without consistent, measurable benefits in language and development.

There are no clear risk factors that identify children who should have PE tubes placed. Some evidence suggests that children with poor baseline hearing (i.e., >25 dB) and those in daycare obtain more of a hearing benefit from PE tubes. In addition, there is limited evidence that children with baseline language or other developmental delays and hearing loss may benefit from earlier PE tube placement.

Overall, the literature suggests that watchful waiting for at least three months is an appropriate initial step in the management of OME. The literature is less clear on management following this initial three months, with some evidence suggesting that even waiting as long as six months may not have deleterious effects on language and development in many children. In terms of other treatment options, there is no evidence that antihistamines, decongestants or nasal steroids are effective treatments for OME.

Adenoidectomy may improve middle ear effusions at six months but does not lead to significant improvements in hearing or in recurrent acute otitis media. Autoinflation may have some benefits in terms of resolution of effusion but may be difficult to use in young patients who might not be cooperative with the treatment. Oral steroids show short-term benefits for OME but fail to sustain these improvements over the longer term. Oral antibiotics may also improve OME in the short term, but the low quality of the evidence does not allow for definitive conclusions. Prophylactic antibiotics are also modestly effective at decreasing the number of episodes of acute otitis media in children with recurrent disease. There is concern for the development of antibiotic resistance with their chronic use, and despite the modest benefits, their use for recurrent acute otitis media and OME has declined.

Guidelines

Two guidelines that address the surgical management of OME (a joint guideline produced by the American Academy of Family Physicians, American Academy of Otolaryngology – Head and Neck Surgery, and American Academy of Pediatrics [AAFP/AAOHNS/AAP]; a National Institute for Health and Clinical Excellence [NICE] guideline produced by the National Collaborating Centre for Women's and Children's

DRAFT

Health) provide similar but slightly different recommendations regarding the management of children with OME. Both recommend monitoring children for the first three months of middle ear effusion and evaluating the child's hearing if the effusion remains at three months. However, NICE recommends hearing testing both at the time of initial diagnosis, and after three months, while the AAFP/ AAOHNS/AAP guideline recommends hearing testing only after OME has been present for three months, unless there is language delay, learning problems or hearing loss is suspected. In addition, language testing is recommended for any child with a documented hearing loss by the AAFP/ AAOHNS/AAP guideline, but not mentioned by the NICE guideline. In addressing this, the text of the evidence review states the following: "A proportion of children referred with suspected OME will also have underlying sensorineural or permanent conductive hearing loss. The GDG [Guideline Development Group] wished to emphasize the need to identify any such component."

Regarding surgical management, the NICE guideline suggests that any child with persistent OME at three months who has a hearing threshold worse than 25 dB should be referred for PE tubes, and if tubes are contraindicated or not desired, then the child should be offered hearing aids and other educational/behavioral interventions. They note that surgical intervention for some children at hearing loss less than 25 to 30 dB may be considered if hearing loss would be expected to significantly impact behavior or development. They specifically identify children with Down syndrome and cleft palate as needing comprehensive specialty care and hearing evaluation, but do not make specific recommendations regarding the timing or use of PE tubes. With regard to the hearing loss level, the text of the evidence review states the following: "Persistent and/or fluctuating OME, resulting in a hearing loss of 25–30 dBHL or greater may have adverse effects on a child's speech and language development, behaviour, emotional development and school progress. *This 25–30 dBHL value is of necessity somewhat notional.* (italics added) Hearing levels fluctuate with time and would not predict the impact precisely even if the hearing history over time were known, because of differing susceptibilities."

In contrast, the AAFP/ AAOHNS/AAP guideline recommends a risk-based approach, in which children at risk for or with language or other developmental delay should be referred more promptly for PE tubes. In children at low risk for delays, the guidelines recommend watchful waiting and monitoring every three to six months until the effusion disappears and referral if significant hearing loss develops or if language or other developmental delays appear. They divide hearing loss into three classes with different actions recommended for each level:

DRAFT

Hearing Level	Recommended Action
≥ 40 dB (moderate hearing loss)	Comprehensive audiologic exam and if hearing loss persists at this level, surgery recommended.
21-39 dB (mild hearing loss)	Comprehensive audiologic exam. Individualize based on effusion duration, severity of hearing loss, parent/caregiver preference: can include optimizing listening and learning environment. Repeat hearing testing in 3-6 months if otitis media with effusion persists and tympanostomy tubes have not been placed.
≤ 20 dB (normal hearing)	Repeat hearing test in 3-6 months if otitis media with effusion persists.

The guideline states this recommendation is based on RCTs and observational studies, with a preponderance of benefit over harm. However, specific citations are not provided that pertain directly to the hearing levels noted above. The text of the guideline does provide citations for the following:

"Asymptomatic OME usually resolves spontaneously, but resolution rates decrease the longer the effusion has been present and relapse is common. Risk factors that make spontaneous resolution less likely include:

- Onset of OME in the summer or fall season,
- Hearing loss more than 30-dB HL in the better hearing ear,
- History of prior tympanostomy tubes, and
- Not having had an adenoidectomy."

Overall Summary

Pressure equalization tubes likely decrease the duration of middle ear effusion over the first year. They also provide short-term improvement in hearing that dissipates by 12 months, resulting in no long-term benefits in language and development as a result of PE tube placement for OME. There are no clear risk factors that identify children who should have PE tubes placed. Some evidence suggests that children with poor baseline hearing (i.e., >25 dB) obtain more of a hearing benefit from PE tubes. Watchful waiting for at least three months and possibly up to six is an appropriate initial step in the management of OME. There is no evidence that antihistamines, decongestants or nasal steroids are effective treatments for OME. Adenoidectomy may improve middle ear effusions at six months but does not lead to significant improvements in hearing or in recurrent acute otitis media. Autoinflation may have some benefits in terms of resolution of effusion, while oral steroids and antibiotics show short-term benefit for OME, but longer term improvement is either not sustained or is uncertain. Prophylactic antibiotics

DRAFT

modestly decrease the number of episodes of acute otitis media in children with recurrent disease.

PROCEDURE

Placement of pressure equalization tubes
Pharmacotherapy
Autoinsufflation

DIAGNOSES

Acute otitis media
Chronic otitis media with effusion

APPLICABLE CODES

CODES	DESCRIPTION
ICD-9 Diagnosis Codes	
381	Nonsuppurative otitis media and eustachian tube disorder
382	Suppurative and unspecified otitis media
315.34	Speech and language developmental delay due to hearing loss
389	Hearing loss
ICD-9 Volume 3 (Procedure Codes)	
None	
CPT Codes	
42820	Tonsillectomy and adenoidectomy; younger than age 12
42821	Tonsillectomy and adenoidectomy; age 12 and over
42830	Adenoidectomy, primary; younger than age 12
42831	Adenoidectomy, primary; age 12 and over
42835	Adenoidectomy, secondary; younger than age 12
42836	Adenoidectomy, secondary; age 12 and over
69433	Tympanostomy (requiring insertion of ventilating tube, local or topical anesthesia)
69436	Tympanostomy (requiring insertion of ventilating tube, general anesthesia)
69424	Ventilating tube removal requiring general anesthesia
HCPCS Codes	
None	

Note: Inclusion on this list does not guarantee coverage

DRAFT

Coverage guidance is prepared by the Health Evidence Review Commission (HERC), HERC staff, and subcommittee members. The evidence summary is prepared by the Center for Evidence-based Policy at Oregon Health & Science University (the Center). This document is intended to guide public and private purchasers in Oregon in making informed decisions about health care services.

The Center is not engaged in rendering any clinical, legal, business or other professional advice. The statements in this document do not represent official policy positions of the Center. Researchers involved in preparing this document have no affiliations or financial involvement that conflict with material presented in this document.

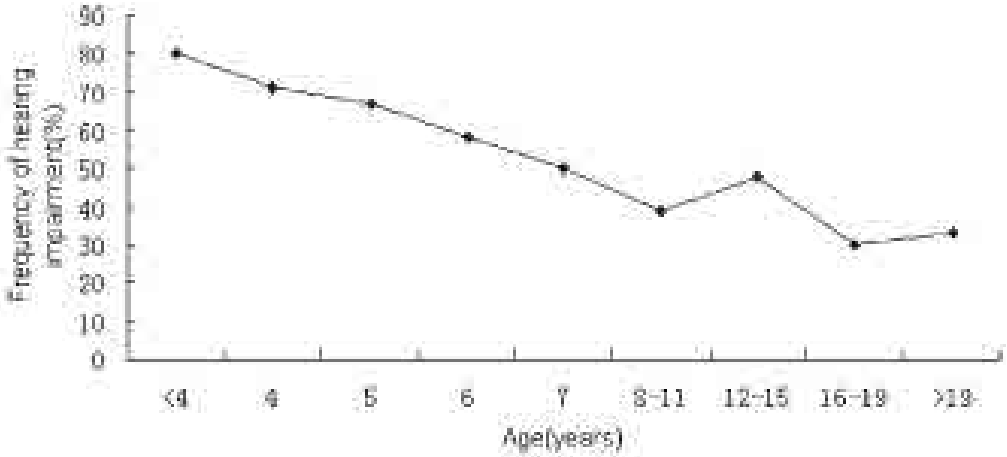
DRAFT

HERC Coverage Guidance – Management of Chronic Otitis Media with Effusion in Children
Disposition of Public Comments

General Comments

Stakeholder	#	Comment	Disposition
OHP Managed Care Medical Directors Oregon	1	There is a lack of clarity about how long the hearing loss needs to be present. Adding the word "persistent" to the second paragraph, "and if documented PERSISTENT hearing loss is greater than or equal to 25dB" allows for clarification that the hearing loss is not just at the onset of an acute otitis media but truly persists.	Thank you for your comment. See change to guidance box to increase clarity.
	2	This therapy is still not effective at changing long term outcomes, yet the guidance still recommends coverage after 3-6 month waiting period. This should be clarified further in the box. OHP does not have tympanostomy tubes in the funded region because of lack of evidence of significant benefit. It would be helpful to have this lack of efficacy up front in the box.	See change in guidance box pertaining to referral for surgery. Changed from "should" to "may".
Chief, Pediatric Otolaryngology Portland, OR	3	I appreciate the opportunity to have a voice and make comments regarding the guidelines for the treatment of otitis media in children. I think the current guidelines are well crafted with excellent documentation of its evidenced-based approach to the treatment of otitis media in children. Toward the end of the guidelines there is a comment about children with cleft palate[1], Down syndrome[2], and other problems where there is a known high incidence of eustachian tube dysfunction (ETD). Other conditions or syndrome known to have high incidence of ETD include: craniofacial abnormalities (e.g. Apert syndrome), achondroplasia[3,4], deletion 22 syndrome (DiGeorge, velocardiofacial syndrome). The proposed guidelines state that these children should be treated individually. As a Pediatric Otolaryngologist who frequently takes care of these Children with Medical Complexity, I want to amplify on this statement, as I have frequently encountered resistance to my treatment plan for these children.	Thank you for taking the time to comment.
	4	Children at risk for chronic otitis media and conductive hearing loss frequently present during infancy, and ear tubes are placed using the current OHP criteria. Their primary care provider or the otolaryngologist then sees these children periodically. For example, I commonly see children with Down syndrome or cleft palate about every six months to monitor their ears, hearing, and other head and neck problems. It is common for these children to present at follow up with one tube which has extruded and otitis media has developed (due to the ETD), yet the other tube is in place and functioning. Upon testing their hearing we discover that they have single sided deafness; one ear without the ear tube has a conductive hearing loss, whereas the other ear hears normally. Until the current guidelines it has been my practice, in congruence with AAP guidelines[5] for management of Down syndrome patients, that I would replace both ear tubes. That is, the likelihood for bilateral hearing loss once the other ear tube extrudes is very high in these CMC patients. Under current DMAP guidelines these children are not eligible for ear tube placement because they have normal hearing in at least one ear. I must wait until the other tube falls out before replacement can be authorized. The loss of the ability to anticipate hearing loss means that the child may suffer several months of bilateral conductive hearing loss before they return to me for care. The reality is that there is often no audiology follow-up at a primary care provider with these children. (Tympanometry is not a hearing test.) Often these children do not present with ear pain or other symptoms of ear infection, they simply have a decline in their hearing that is discovered at the next visit to a pediatric audiologist or otolaryngologist. Also unilateral hearing loss can have an important negative impact on school performance, behavior, and speech and language development[6].	Thank you for sharing your professional experience. The EbGS understands that the population you describe is at increased risk for OME and hearing loss, hence the reference to an individualized treatment plan. EbGS believes this will allow coverage on a case by case basis.

DRAFT

Stakeholder	#	Comment	Disposition
	5	I know most of the references that you cited in the development of these proposed guidelines. If one peruses these articles carefully, especially the randomized control trials, you find in the materials and methods section that these high-risk groups were excluded from the studies. That is, the researchers knew that introducing children with Down syndrome, for example, would muddy their data. I would suggest that the criteria proposed, based on the study of otherwise normal children, should not used as treatment guidelines for children with the aforementioned high-risk ETD. There is data that shows, for example, that 7 years is the median age at which a child with a cleft palate develops eustachian tube dysfunction comparable to the age match norms[7,8]. In other words it takes at least seven years for 50% of children with cleft palate to start to have eustachian tube function akin to their playmates in the classroom (see Figure 1). A three-year- old child with cleft palate (or other high-risk ETD disorders) should be allowed to have ear tubes placed once unilateral conductive hearing loss is found upon the first ear tube extruding. I hope you will strengthen the language of your guidelines to help Oregon Otolaryngologists find it easier to obtain authorization for treatment of ear disease in this high-risk population.  **FIGURE 1 Frequency of hearing impairment at varying ages in 508 ears of unrepaired patients.**	EbGS agrees that the data on high-risk groups is very limited. In addition, the evidence is lacking regarding the impact of unilateral hearing loss. To address the uncertainty of the effect of treatment of OME on this population, the EbGS has allowed for an individualized treatment plan in this population. See comment #4.

HEALTH EVIDENCE REVIEW COMMISSION (HERC)

DRAFT COVERAGE GUIDANCE: PERCUTANEOUS INTERVENTIONS FOR LOW BACK PAIN

DATE: XX/XX/XXXX

HERC COVERAGE GUIDANCE

For radicular low back pain, Epidural steroid injections should be covered for patients with persistent radiculopathy due to herniated lumbar disc; it is recommended that shared decision-making regarding epidural steroid injection include a specific discussion about inconsistent evidence showing moderate short-term benefits, and lack of long-term benefits. If an epidural steroid injection does not offer benefit, repeated injections should not be covered.

Epidural steroid injections should NOT be covered for central spinal stenosis.

For radicular low back pain, the following treatments should NOT be covered:

- coblation nuceleoplasty
- radiofrequency denervation

For nonradicular low back pain, the following treatments should NOT be covered:

- facet joint corticosteroid injection
- prolotherapy
- intradiscal corticosteroid injection
- local injections
- botulinum toxin injection
- epidural steroid injection
- intradiscal electrothermal therapy (IDET)
- ~~therapeutic~~ medial branch block
- radiofrequency denervation
- sacroiliac joint steroid injection
- coblation nucleoplasty
- percutaneous intradiscal radiofrequency thermocoagulation

*Coverage guidance for non-pharmacologic interventions, pharmacologic interventions, and imaging for low back pain are addressed in separate documents.

DRAFT

Oregon Health Authority

RATIONALE FOR GUIDANCE DEVELOPMENT

The HERC selects topics for guideline development or technology assessment based on the following principles:

- Represents a significant burden of disease
- Represents important uncertainty with regard to efficacy or harms
- Represents important variation or controversy in clinical care
- Represents high costs, significant economic impact
- Topic is of high public interest

Coverage guidance development follows to translate the evidence review to a policy decision. Coverage guidance may be based on an evidence-based guideline developed by the Evidence-based Guideline Subcommittee or a health technology assessment developed by the Heath Technology Assessment Subcommittee. In addition, coverage guidance may utilize an existing evidence report produced by one of HERC's trusted sources, generally within the last three years.

EVIDENCE SOURCES

Livingston, C., Little, A., King, V., Pettinari, C., Thielke, A., Pensa, M., Vandegriff, S., & Gordon, C. (2012). *State of Oregon Evidence-based Clinical Guidelines Project. Percutaneous interventions for low back pain: A clinical practice guideline based on the 2009 American Pain Society Guideline (Interventional Therapies, Surgery, and Interdisciplinary Rehabilitation for Low Back Pain).* Salem: Office for Oregon Health Policy and Research. Retrieved from http://www.oregon.gov/OHA/OHPR/HERC/Evidence-Based-Guidelines.shtml

Chou, R., Loesser, J.D., Owens, D.K., Rosenquist, R.W., Atlas, S.J., Baisden, J., Carragee, E.J., Grabois, M., Murphy, D.R., Resnick, D.K., Stanos, S.P., Shaffer, W.O., Wall E.M. (2009) Interventional therapies, surgery, and interdisciplinary rehabilitation for low back pain: An evidence-based clinical practice guideline from the American Pain Society. *Spine* 34:10:1066-1077. – accompanied by:

Chou, R., Atlas, S.J., Stanos, S.P., Rosenquist, R.W. (2009). A review of the evidence for an American Pain Society clinical practice guideline. *Spine* 34:10:1078-1094.

The summary of evidence in this document is derived directly from these evidence sources, and portions are extracted verbatim.

DRAFT

SUMMARY OF EVIDENCE

Clinical Background

Low back pain is the fifth most common reason for all physician visits in the United States. Approximately one quarter of US adults reported having low back pain lasting at least one whole day in the past three months, and 7.6% reported at least one episode of severe acute low back pain within a 1-year period. Low back pain is also very costly. Total incremental direct health care costs attributable to low back pain in the US were estimated at $26.3 billion in 1998. In addition, indirect costs related to days lost from work are substantial, with approximately 2% of the US work force compensated for back injuries each year.

Many patients have self-limited episodes of acute low back pain and do not seek medical care. Among those who do seek medical care, pain, disability, and return to work typically improve rapidly in the first month. However, up to one third of patients report persistent back pain of at least moderate intensity one year after an acute episode, and one in five report substantial limitations in activity. Approximately 5% of the people with back pain disability account for 75% of the costs associated with low back pain.

Many options are available for evaluation and management of low back pain. However, there has been little consensus, either within or between specialties, on appropriate clinical evaluation and management of low back pain. Numerous studies show unexplained, large variations in use of diagnostic tests and treatments. Despite wide variations in practice, patients seem to experience broadly similar outcomes, although costs of care can differ substantially among and within specialties.

Evidence Review

Recommendation #1: In patients with persistent radiculopathy due to herniated lumbar disc, it is recommended that clinicians discuss risks and benefits of epidural steroid injection as an option (weak recommendation, moderate-quality evidence). It is recommended that shared decision-making regarding epidural steroid injection include a specific discussion about inconsistent evidence showing moderate short-term benefits, and lack of long-term benefits. There is insufficient evidence to adequately evaluate benefits and harms of epidural steroid injection for spinal stenosis.

For radiculopathy due to herniated lumbar disc, evidence on benefits of epidural steroid injection is mixed. Although some higher-quality trials found epidural steroid injection associated with moderate short-term (through up to 6 weeks) benefits in pain or function, others found no differences *versus* placebo injection. Reasons for the discrepancies between trials is uncertain, but could be related to the type of comparator treatment, as trials that compared an epidural steroid injection to an epidural saline or local anesthetic injection tended to report poorer results than trials that compared an

epidural steroid injection to a soft-tissue (usually interspinous ligament) placebo injection. Regardless of the comparator intervention, there is no convincing evidence that epidural steroids are associated with long-term benefits and most trials found no reduction in rates of subsequent surgery. Although serious complications following epidural steroid injection are rare in clinical trials, there are case reports of paralysis and infections. There is insufficient evidence on clinical outcomes to recommend a specific approach for performing epidural steroid injection, or on use of fluoroscopic guidance. In addition, insufficient evidence exists to recommend how many epidural injections to perform, though one higher-quality trial found that if an initial epidural steroid injection did not result in benefits, additional injections over a 6-week period did not improve outcomes.

Decisions regarding use of epidural steroid injection should be based on a shared decision-making process that includes a discussion of the inconsistent evidence for short-term benefit, lack of long-term benefit, potential risks, and costs. Patient preferences and individual factors should also be considered. For example, epidural steroid injection may be a reasonable option for short-term pain relief in patients who are less optimal surgery candidates due to comorbidities. There is insufficient evidence to guide specific recommendations for timing of epidural steroid injection, though most trials enrolled patients with at least subacute (greater than 4 weeks) symptoms.

Evidence on efficacy of epidural steroid injection for spinal stenosis is sparse and shows no clear benefit, though more trials are needed to clarify effects. Although chymopapain chemonucleolysis is effective for radiculopathy due to herniated lumbar disc, it is less effective than discectomy and is no longer widely available in the United States, in part due to risk of severe allergic reactions. Three trials suggest that intradiscal steroid injection has similar efficacy to chemonucleolysis, although none were placebo controlled.

Recommendation #2: In patients with persistent nonradicular low back pain, facet joint corticosteroid injection, prolotherapy, and intradiscal corticosteroid injection are not recommended (strong recommendation, moderate-quality evidence).

Injections and most interventional therapies for nonradicular low back pain target specific areas of the back that are potential sources of pain, including the muscles and soft tissues (botulinum toxin injection, prolotherapy, and local injections), facet joints (facet joint steroid injection, therapeutic medial branch block, and radiofrequency denervation), degenerated intervertebral discs (intradiscal steroid injection, IDET, and related procedures), and sacroiliac joints (sacroiliac joint injection). There is no convincing evidence from randomized trials that injections and other interventional therapies are effective for nonradicular low back pain. Facet joint steroid injection,

prolotherapy and intradiscal steroid injections are not recommended because randomized trials consistently found them to be no more effective than sham therapies.

Recommendation #3: There is insufficient evidence to adequately evaluate benefits of local injections, botulinum toxin injection, epidural steroid injection, intradiscal electrothermal therapy (IDET), therapeutic medial branch block, radiofrequency denervation, sacroiliac joint steroid injection, coblation nucleoplasty, percutaneous intradiscal radiofrequency thermocoagulation …. or other medications for nonradicular low back pain.

For local injections[1], there is insufficient evidence to accurately judge benefits because available trials are small, lower-quality, and evaluate heterogeneous populations and interventions. Trials of IDET and radiofrequency denervation reported inconsistent results. There were a small number of higher quality trials, and in the case of radiofrequency denervation, the trials had technical or methodologic shortcomings, making it difficult to reach conclusions about benefits**.** For other interventional therapies, data are limited to one to two small placebo-controlled randomized trials (botulinum toxin injection, epidural steroid injection for nonradicular low back pain, PIRFT and sacroiliac joint steroid injection), or there are no placebo-controlled randomized trials (therapeutic medial branch block, coblation nucleoplasty….or other medications).

[Evidence Source]

Overall Summary

For radiculopathy due to herniated lumbar disc, evidence on benefits of epidural steroid injection is mixed, with some trials finding moderate short-term benefits and others finding no differences. There is no convincing evidence that epidural steroids are associated with long-term benefits and most trials found no reduction in rates of subsequent surgery. For nonradicular low back pain, there is likewise no convincing evidence that injections and other interventional therapies are effective, while there is consistent evidence that facet joint steroid injection, prolotherapy and intradiscal steroid injections are no more effective than sham therapies.

PROCEDURE

Epidural steroid injection
Botulinum toxin injection
Local injections
Facet joint steroid injection

[1] Defined as placement of a local anesthetic into the muscles or soft tissues of the back via a catheter. One type of local injection is trigger point injection.

DRAFT

Therapeutic medial branch block
Radiofrequency denervation
Intradiscal steroid injection
Intradiscal electrothermal therapy (IDET)
Sacroiliac joint injection
Chymopapain chemonucleolysis
Coblation nucleoplasty
Percutaneous intradiscal radiofrequency thermocoagulation (PIRFT)

DIAGNOSES

Low back pain

APPLICABLE CODES

CODES	DESCRIPTION
ICD-9 Diagnosis Codes	
720.1	Spinal enthesopathy
720.2	Sacroiliitis, not elsewhere classified
721.3	Lumbosacral spondylosis without myelopathy
721.42	Spondylosis with myelopathy, lumbar region
721.5	Kissing spine
721.6	Ankylosing vertebral hyperostosis
721.7	Traumatic spondylopathy
721.8	Other allied disorders of spine
721.9	Spondylosis of unspecified site
722.1	Displacement of thoracic or lumbar intervertebral disc without myelopathy
722.2	Displacement of intervertebral disc, site unspecified, without myelopathy
722.32	Schmorl's nodes, lumbar region
722.39	Schmorl's nodes, other region
722.5	Degeneration of thoracic or lumbar intervertebral disc
722.6	Degeneration of intervertebral disc, site unspecified
722.70	Intervertebral disc disorder with myelopathy, unspecified region
722.72	Intervertebral disc disorder with myelopathy, thoracic region
722.73	Intervertebral disc disorder with myelopathy, lumbar region
722.80	Postlaminectomy syndrome, unspecified region
722.82	Postlaminectomy syndrome, thoracic region
722.83	Postlaminectomy syndrome, lumbar region
722.90	Other and unspecified disc disorder, unspecified region
722.92	Other and unspecified disc disorder, thoracic region
722.93	Other and unspecified disc disorder, lumbar region
724	Other and unspecified disorders of back
724.0	Spinal stenosis other than cervical
724.00	Spinal stenosis, unspecified region
724.01	Spinal stenosis, thoracic region
724.02	Spinal stenosis, lumbar region, without neurogenic claudication
724.03	Spinal stenosis, lumbar region, with neurogenic claudication
724.09	Spinal stenosis, other region
724.1	Pain in thoracic spine

DRAFT

CODES	DESCRIPTION
724.2	Lumbago
724.3	Sciatica
724.4	Thoracic or lumbosacral neuritis or radiculitis, unspecified
724.5	Backache, unspecified
724.6	Disorders of sacrum
724.7	Disorders of coccyx
724.70	Unspecified disorder of coccyx
724.71	Hypermobility of coccyx
724.79	Other disorders of coccyx
724.8	Other symptoms referable to back
724.9	Other unspecified back disorders
730.2	Unspecified osteomyelitis
732.0	Juvenile osteochondrosis of spine
733.0	Osteoporosis
737.2	Lordosis (acquired)
737.30	Scoliosis [and kyphoscoliosis], idiopathic
737.39	Other kyphoscoliosis and scoliosis
737.4	Curvature of spine associated with other conditions
737.8	Other curvatures of spine
737.9	Unspecified curvature of spine
738.4	Acquired spondylolisthesis
738.5	Other acquired deformity of back or spine
739.2	Nonallopathic lesions, thoracic region
739.3	Nonallopathic lesions, lumbar region
739.4	Nonallopathic lesions, sacral region
754.2	Congenital musculoskeletal deformities of spine
756.1	Congenital anomalies of spine
846	Sprains and strains of sacroiliac region
847.1	Sprain of thoracic
847.2	Sprain of lumbar
847.3	Sprain of sacrum
847.4	Sprain of coccyx
847.9	Sprain of unspecified site of back
ICD-9 Volume 3 (procedure codes)	
87.24	Other x-ray of lumbosacral spine
88.38	Other computerized axial tomography
88.93	X-ray, other and unspecified
CPT	
0216T	Injection(s), diagnostic or therapeutic agent, paravertebral facet (zygapophyseal) joint (or nerves innervating that joint) with ultrasound guidance; lumbar or sacral, single level
0217T	second level
0218T	third and any additional level(s)
20552	Injection, single or multiple trigger point(s), 1 or 2 muscle(s)
20553	Injection, single or multiple trigger point(s), 3 or more muscle(s)
20600	Arthrocentesis, aspiration and /or injection; small joint or bursa (eg, fingers, toes)
20605	intermediate joint or bursa (eg, temporomandibular, acromioclavicular, wrist, elbow or ankle, olecranon bursa)
20610	major joint or bursa (eg, shoulder, hip, knee joint, subacromial bursa)
22526	Percutaneous intradiscal electrothermal annuloplasty, unilateral or bilateral including

DRAFT

CODES	DESCRIPTION
	fluoroscopic guidance; single level
22527	1 or more additional levels
27096	Injection procedure for sacroiliac joint, anesthetic steroid, with image guidance (fluoroscopy or CT) including arthrography when performed
62292	Injection procedure, arterial, for occlusion of arteriovenous malformation, spinal
64412	Injection, anesthetic agent; spinal accessory nerve
64483	Injection(s), anesthetic agent and/or steroid, transforaminal epidural, with imaging guidance (fluoroscopy or CT); lumbar or sacral, single level
64484	Injection(s), anesthetic agent and/or steroid, transforaminal epidural, with imaging guidance (fluoroscopy or CT); lumbar or sacral, each additional level
64493	Injection(s), diagnostic or therapeutic agent, paravertebral facet (zygapophyseal) joint (or nerves innervating that joint) with image guidance (fluoroscopy or CT); lumbar or sacral, single level
64494	second level
64495	third and any additional level(s)
64635	Destruction by neurolytic agent, paravertebral facet join nerve(s), with imaging guidance (fluoroscopy or CT); lumbar or sacral, single facet joint
64636	Destruction by neurolytic agent, paravertebral facet join nerve(s), with imaging guidance (fluoroscopy or CT); lumbar or sacral, each additional facet joint
76942	Ultrasonic guidance for needle placement (eg, biopsy, aspiration, injection, localization device), imaging supervision and interpretation
77002	Fluoroscopic guidance for needle placement (eg, biopsy, aspiration, injection, localization device), imaging supervision and interpretation
77003	Fluoroscopic guidance and localization of needle or catheter tip for spine or paraspinous diagnostic or therapeutic injection procedures (epidural or subarachnoid)
77021	Magnetic resonance guidance for needle placement (eg, for biopsy, needle aspiration, injection, or placement of localization device) radiological supervision and interpretation
96372	Therapeutic, prophylactic, or diagnostic injection (specify substance or drug); subcutaneous or intramuscular
HCPCS Codes	
M0076	Prolotherapy
S2348	Decompression procedure, percutaneous, of nucleus pulposus of intervertebral disc, using radiofrequency energy, single or multiple levels, lumbar

Note: Inclusion on this list does not guarantee coverage

Coverage guidance is prepared by the Health Evidence Review Commission (HERC), HERC staff, and subcommittee members. The evidence summary is prepared by the Center for Evidence-based Policy at Oregon Health & Science University (the Center). This document is intended to guide public and private purchasers in Oregon in making informed decisions about health care services.

The Center is not engaged in rendering any clinical, legal, business or other professional advice. The statements in this document do not represent official policy positions of the Center. Researchers involved in preparing this document have no affiliations or financial involvement that conflict with material presented in this document.

HERC Coverage Guidance – Percutaneous Interventions for Low Back Pain
Disposition of Public Comments

General Comments

Stakeholder	#	Comment	Disposition
OHP Managed Care Medical Directors Oregon	1	"Define radiculopathy, and persistent radiculopathy in particular. Are their subgroups who benefit more or less from epidural steroid injection.	The evidence source included a total of 20 placebo controlled trials of epidural steroid injections that had mixed results. All populations were described as having radiculopathy or neurogenic claudication, with some requiring imaging evidence of a prolapsed disc. Regarding how radiculopathy or neurogenic claudication is defined: Of the five trials reviewed, none required objective findings on physical exam other than positive straight leg raising, and only two required imaging confirmation. The EbGS regrets that they are unable to provide more definition.
	2	Are there subgroups of folks with radiculopathy who do better, compared to those who don't (i.e. for epidural steroid injections)?	The evidence source does not report on this, other than those trials for which the comparator was epidural saline or anaesthetic were more likely to be negative than those trials for which the comparator was a soft tissue injection.
Physician, Pain Management Eugene, OR	3	The proposed coverage guidance for percutaneous interventions for low back pain fails to serve the best interest of the patient in pain. The recommendations seem to be based on biased clinical practice guidelines and not on primary published literature. As these "literature reviews" form the basis for your recommendations, I will refer to this type of analysis as well.	The evidence source is the American Pain Society guideline on interventional therapies for LBP, which is supported by a systematic review of the evidence. This review includes the "primary published literature". It is not clear why the commenter believes the APS guidelines are biased.
	4	The inclusion of a discussion about 'shared decision-making' in treatment recommendations is redundant and unnecessary. It seems the Commission recommendations for a procedure are not equal in strength to those against a procedure. If the commission is willing to support non-coverage of a procedure then they should also be willing to support coverage of a procedure. I note that the 'shared decision making' discussion is not (as it should be) included for non-covered procedures.	The EbGS disagrees that shared decision making is unnecessary when the evidence supporting the effectiveness of an intervention is conflicting. The patient should understand the lack of certainty regarding effectiveness, particularly when there are potential harms. When there is evidence that a treatment is ineffective, the treatment should not be offered, hence there is no need for shared decision making.
	5	Research into pain treatment is difficult; one of the most obvious is the absence of an objective verifiable measure of the experience of pain. Such a measure would allow for greater power from smaller studies. A type two error of publishing a negative result when a difference does not exist is particularly at risk of occurring in these studies. Very few trials of pain treatments are adequately powered to prevent type two errors.	There are numerous scales used to assess pain, and many studies have shown significant results, including in this evidence base. Seven of the 20 trials of epidural steroids included in the systematic review used validated pain measures and found positive results.
	6	I address specific recommendations in the order of your document.	In the referenced systematic review (Parr 2012), there were

DRAFT

Stakeholder	#	Comment	Disposition
		1. Radicular low back pain. Extensive trials of moderate quality confirm a meaningful clinical response to transforaminal epidural injections. Good evidence also exists for the use of caudal epidural injections in the management of lumbar discogenic pain, spinal stenosis and post surgery syndrome. (Parr 2012) Based on this analysis, exclusion of spinal stenosis is not appropriate.	3 studies that addressed spinal stenosis, only one of which was a RCT. In this trial, the control comparator was bupivacaine injection, and the control group had a slightly better response than the steroid group in pain and functional status. There was no non-injection control group. The other two studies were case series.
	7	2. Nonradicular low back pain. Extensive level II evidence exists for lumbar and cervical diagnostic and therapeutic facet denervation procedures as well as caudal epidural injection of local anesthetic and steroids. (Manchikanti 2009)	The search dates for the stated reference are not specified, but publication date was 2009 (same as the Chou review). Without search dates, the EbGS is unable to evaluate the currency of the literature review. In addition, the guideline was reviewed at the time that the Oregon Guideline Development Group was developing the "Percutaneous Interventions for Low Back Pain Clinical Practice Guideline" and was rated poor quality.
	8	3. 'Local injection' is not adequately specific to include in your non covered list. If you intend to exclude 'trigger point injections' then you should state that and offer evidence, as there is an extensive literature on that technique.	Local injections in the evidence review are defined as placement of a local anesthetic into the muscles or soft tissues of the back via a catheter. One type of local injection is trigger point injection. *Clarification added to the guidance document.* The evidence review found insufficient evidence to draw conclusions for any type of local injection. Of the 4 trials identified, two addressed trigger point injections. None had follow up longer than 2 weeks, and the trial sizes ranged from 15 to 63.
North American Spine Society (NASS) Burr Ridge, IL	9	The North American Spine Society (NASS) wishes to comment on the Evidence Review Commission (HERC) draft coverage guidance regarding percutaneous interventions for low back pain (LBP), specifically regarding the non-coverage recommendation for epidural steroid injections (ESI) in the treatment of radicular pain due to spinal stenosis, in addition to sacroiliac injections and radiofrequency denervation for non-radicular LBP. NASS is a multispecialty medical organization dedicated to fostering the highest quality, evidence-based, ethical spine care. NASS has over 5,500 members from several disciplines, including orthopedic surgery, neurosurgery, physiatry, pain management and other spine professionals.	Thank you for your comment.
	10	Your policy states that there is insufficient evidence to adequately evaluate benefits and harms of ESIs for spinal stenosis. We strongly disagree with this assessment and present information to assist in a balanced review. Unfortunately, your HERC draft only sited three	The APS guideline is based on a full systematic review of the evidence.

DRAFT

Stakeholder	#	Comment	Disposition
		references of which all are based on the controversial and criticized 2009 practice guidelines published by the American Pain Society.	
	11	Various organizations and professional medical societies have addressed the efficacy of ESIs for the treatment of radicular pain. Your draft actually states "ESIs should be covered for patients with persistent radiculopathy due to herniated lumbar disc", however you should realize, herniated discs often cause stenosis contributing to the radicular symptoms which the HERC draft recommends coverage for.	The EbGS is aware that lumbar disc herniation may contribute to central spinal stenosis, but is also aware that other mechanical factors such as facet hypertrophy and spondylosis contribute as well. The presentation of spinal stenosis is distinct from that of radiculopathy due to herniated disc. Studies included in the evidence review generally included patients with unilateral sciatica with signs for nerve root irritation or compression. They did not include patients with the typical presentation of spinal stenosis (neurogenic claudication, often bilateral, relieved by change in posture). Spinal stenosis involves compression at the level of the spinal cord, while radiculopathy refers to compression of the nerve root. *To clarify this distinction, "central" was added to the guidance document.*
	12	Recently, the Washington State HTA committee voted to cover therapeutic lumbar (and cervical/thoracic) ESIs for the treatment of radicular pains due to stenosis and other etiologies. Coverage was supported and endorsed by 11 national and international specialty societies[1]. Their decisions were based on an extensive, rigorous review, which took over a year to complete. The committee concluded, "The current evidence on spinal injections demonstrates that there is sufficient evidence to cover with conditions the use of therapeutic ESIs in the lumbar or cervical-thoracic spine for chronic pain". They further stated "Based on the evidence about technologies', safety, efficacy, and cost-effectiveness, therapeutic ESIs in the lumbar or cervical-thoracic spine is a covered benefit when all of the following conditions are met: 1. For treatment of radicular pain 2. With fluoroscopic guidance or CT guidance 3. After failure of conservative therapy 4. No more than two without clinically meaningful improvement in pain and function 5. Maximum of 3 in 6 months."	The EbGS disagrees that the current recommended guidance differs significantly from the WA HTA decision. This report specifically states that there is NO benefit of epidural steroid injections in either the short or long term in patients with spinal stenosis. The WA HTA does indeed recommend coverage of ESIs for patients with radicular pain, but does not specify the etiology of the radicular pain, nor do they state "radicular pains due to stenosis".
	13	Additionally, NASS has published guidelines on degenerative spinal stenosis, which concluded there is evidence (Level II-III studies) demonstrating good short-term efficacy of	The reference cited is a guideline published in 2007, before the date of the Chou evidence review (2009), which also

DRAFT

Stakeholder	#	Comment	Disposition
		fluoroscopically-guided transforaminal ESIs in the treatment of radicular pain due to stenosis[2]. The guidelines also concluded that a multiple injection protocol of fluoroscopically-guided transforaminal ESIs can produce long-term relief of pain in patients with radiculopathy or intermittent neurogenic claudication due to lumbar spinal stenosis. The role of ESIs in the treatment of lumbar spinal stenosis was one question the NASS guideline addressed.	addressed ESIs for lumbar spinal stenosis.
	14	Other societies, including the American Society of Interventional Pain Physicians (ASIPP) have guidelines and reviews that evaluated the efficacy of epidurals for spine and radicular pain and concluded that the evidence for transforaminal epidurals is strong for short-term and moderate for long-term improvement in managing chronic LBP and sciatica[3].	The reference cited is a guideline published in 2007, before the date of the Chou evidence review (2009), which also addressed ESIs for chronic LBP.
	15	There have been a number of randomized, prospective studies indicating good short and long-term efficacy of fluoroscopically-guided transforaminal ESIs in the treatment of radicular pain[4-9]. One such prospective, randomized, controlled, double-blinded study of 55 patients with > 6 weeks radicular pain with radiographically confirmed nerve root compression due to either herniated nucleus pulposus or foraminal stenosis, were treated with 1-3 fluoroscopically-guided transforaminal ESIs of bupivacaine, with or without steroid[6]. All were candidates and agreed to proceed with surgery. Follow-up was 13-26 months and over a mean of 23 months from injection, 71 percent of patients in treatment group cancelled surgery vs. 33 percent control group ($p<.004$). The study concluded that selective nerve root blocks (transforaminal epidural injections) with corticosteroids are more effective than bupivacaine alone, obviating the need for decompressive surgery and all patients with 1 or 2 level radiculopathy should be treated with selective nerve root blocks prior to considering surgery.	The EbGS does not disagree that ESIs can be effective for radiculopathy due to disc herniation, hence the permissive coverage guidance. In addition, the citations listed were published before the date of the Chou evidence review. The EbGS bases their guidance documents on reviews of the literature that utilize the highest standards of evidence based medicine. Studies are included or excluded based on transparent, reproducible criteria; therefore the EbGS does not investigate individual studies. The EbGS assumes that the conclusions reached by the authors of these reviews weigh all the available evidence in accordance with the principles of evidence based medicine, and does not attempt to re-review the entire body of evidence to reach its own conclusions.
	16	A recent systematic review of ESIs in the treatment of chronic LBP concluded that a fluoroscopically-guided ESI is a reasonable treatment option in patients unresponsive to physical therapy and anti-inflammatory medications[10].	The reference cited was published in 2008, before the date of the Chou evidence review (2009).
	17	This policy also does not address the diagnostic aspect of selective spinal nerve blocks that utilize the same CPT codes (64479-64484) as therapeutic transforaminal epidural injections. Anesthetizing a specific spinal nerve and subsequently measuring the amount of radicular pain relief can help provide diagnostic confirmation as to the source of a patient's limb (radicular) pain and which nerve root may need to be surgically decompressed in patients with radicular pain due to central and/or foraminal spinal stenosis. Diagnostic spinal nerve blocks are not only highly sensitive and specific at diagnosing radicular pain, but may be superior to imaging studies (e.g., MRI) alone in	The Chou review addressed this intervention; their comments are as follows: "because nerve root compression can usually be identified by non-invasive imaging, the main roles of diagnostic nerve root blocks are to evaluate the appropriate target level for interventions when multiple nerve roots are involved or to confirm radiculopathy when imaging is equivocal or when there is discordance between clinical findings and imaging. No reliable reference standard

DRAFT

Stakeholder	#	Comment	Disposition
		predicting surgical outcomes[11].	(such as electrophysiologic testing) is available for estimating diagnostic accuracy of selective nerve root blocks for identifying "true" nerve root pain in these situations. We therefore focused our review on evidence on whether use of selective nerve root blocks to select patients for procedures intended to relieve nerve root compression improves clinical outcomes compared to not using selective nerve root blocks to select patients. We identified one lower-quality systematic review on diagnostic accuracy of selective nerve root blocks. However, it included no studies that evaluated whether use of diagnostic selective nerve root blocks to identify patients for procedures intended to relieve nerve root compression improves clinical outcomes compared to relying only on imaging or other non-invasive diagnostic methods to select patients. From 381 potentially relevant citations, we identified no relevant studies." The citation referenced was published before the date of the Chou review; see comment #15.
	18	The main purpose of epidural injections in the treatment of radicular pain due to spinal stenosis is to reduce pain and morbidity while improving function and resumption of normal activities.	The evidence demonstrates that ESIs may be effective for radicular pain. It is not effective for central spinal stenosis.
	19	With respect to radiofrequency neurolysis (RFN) in the treatment of LBP secondary to facet-mediated pain, NASS disagrees with the HERC non-coverage decision. One study of LBP patients with facet-mediated pain diagnosed by controlled medial branch blocks, RFN demonstrated significant pain relief for at least 12 months in a majority of patients[12]. A comprehensive review by Manchikanti et al gave a 1C/strong recommendation for RFN for the treatment of facet-mediated LBP[13]. A recent review by Bogduk, et al demonstrated that lumbar RFN had positive effects on pain and disability and that all valid, randomized controlled trials showed RFN to be more effective than sham treatment[14].	Reference #12 was published before the date of the Chou review. Reference #13 was published in 2009, but does not list search dates for their evidence review. Without search dates, the EbGS is unable to evaluate the currency of the cited review. Reference #14 is a narrative review, no systematic search of the literature was undertaken.
	20	We strongly disagree with the exclusion of sacroiliac joint injections (SIJ) in the evaluation of LBP. Eleven specialty societies endorsed a statement on sacroiliac injections and	The Chou review identified only 1 trial that met inclusion criteria that evaluated SI joint injections. While there was

DRAFT

Stakeholder	#	Comment	Disposition
		submitted it to the Washington HTA as part of a technology review[1]. The Washington HTA committee voted to cover sacroiliac intra-articular injections. We hope that this information will assist in formation of a reasonable coverage policy.	improvement in median pain scores at one month, it included only 24 patients. This was the same evidence included in the WA HTA technology review.
International Spine Intervention Society (ISIS) San Rafael, California	21	The International Spine Intervention Society (ISIS), a multi-specialty association of 3,000 physicians dedicated to the development and promotion of the highest standards for the practice of interventional procedures in the diagnosis and treatment of spine pain would like to comment on the Draft Coverage Guidance for Percutaneous Interventions for Low Back Pain. We commend the Oregon Health Evidence Review Commission (HERC) on incorporating a number of the ISIS guidelines in the draft coverage guidance; however, we would like to bring attention to several issues of concern:	Thank you for your comment.
	22	1. Sacroiliac Joint Steroid Injections for Non-radicular Pain We strongly disagree with the exclusion of sacroiliac joint injections for non-radicular pain and are enclosing information to assist in a balanced and diligent review. 11 national medical specialty societies* endorsed a statement on SI joint injections[1] (see attachment) and submitted it to the Washington State Health Technology Assessment (HTA) Program, as part of a technology review. Upon extensive review, the WA HTA committee voted to cover sacroiliac intra-articular injections of steroids.[2]	See comment #20. The clinical committee of the WA HTA elected to make a coverage decision based on one trial of only 24 patients.
	23	2. Radiofrequency Denervation for Non-radicular Low Back Pain We strongly disagree with the proposal to exclude from coverage radiofrequency (RF) denervation for non-radicular low back pain. Diagnostic lumbar medial branch blocks, which are allowed under this policy, are concordant for lumbar facet join mediated pain. Hence, RF neurotomy should, naturally, also be covered. Other treatment options, such as surgery, are certainly much more invasive and costly.	It is not clear why the commenter believes that diagnostic lumbar medial branch blocks for non-radicular LBP are allowed under this guidance. *The word "therapeutic has been deleted from medial branch block in guidance box.*
	24	The first controlled study, which clearly showed that RF neurotomy was not a placebo, in the lumbar region, was by Van Kleef,[3] though its limited effect was attributed to the sub-optimal placement (not parallel to the nerves) of the electrode. A subsequent yearlong study[4] of carefully screened patients diagnosed as having lumbar zygapophyseal joint pain who underwent denervation by RF neurotomy in which the electrodes were correctly placed parallel to the target nerves, found that, after one year, some 60% of patients can expect to have at least 80% relief of their pain and 80% of patients can expect more than 60% relief. Therefore, this study showed that for patients with lumbar zygapophyseal joint pain, diagnosed by controlled medial branch blocks, lumbar radiofrequency medial branch neurotomy offers a good chance of obtaining worthwhile relief of pain sustained for at least 12 months.	The citations were published before the date of the Chou review. The EbGS bases their guidance documents on reviews of the literature that utilize the highest standards of evidence based medicine. Studies are included or excluded based on transparent, reproducible criteria; therefore the EbGS does not investigate individual studies. The EbGS assumes that the conclusions reached by the authors of these reviews weigh all the available evidence in accordance with the principles of evidence based medicine, and does not attempt to re-review the entire body of evidence to reach its own conclusions.

Stakeholder	#	Comment	Disposition
	25	In addition, Manchikanti et al[5] performed a comprehensive review of therapeutic interventions in managing chronic spinal pain and determined that, for lumbar RF, based on criteria by Guyatt et al.,[6] the recommendation is 1C (strong recommendation).	This is the same review cited by the prior commenter. This review was rated poor quality when evaluated during the development of the Oregon "Percutaneous Interventions for Low Back Pain Clinical Practice Guideline". See comment #19.
	26	3. Epidural Injections for Spinal Stenosis We encourage HERC to reconsider the proposal not to cover Epidural Steroid Injections (ESIs) for spinal stenosis. Studies evaluation the efficacy ESIs for LSS without fluoroscopic guidance should not be considered, although all showed a short-term benefit ranging from 1 week to 2 months of relief, and one demonstrated a longer term benefit with up to 10 months of relief.[7,8,9,10] The more recent studies, which used fluoroscopic guidance,[11,12,13,14] all demonstrated some short-term benefit; while the Botwin study[12], the only prospective evaluation, showed a substantial long-term benefit as well. This study evaluated 34 patients with unilateral radicular symptoms secondary to LSS with fluoroscopically guided transforaminal ESIs. A mean of 1.9 injections per patient was performed. Subjects were evaluated at 2 and 12 months when compared to the pre-injection baseline. The patient satisfaction scale revealed that 62% of patients at 2 months and 64% of patients at 12 months felt somewhat or completely better. Briggs et al.[15], in an evaluation of injection treatment in lumbar spinal stenosis in older adults, reported significant alleviation of pain after injection treatment under fluoroscopy.	See comments # 6 and #24. The only citation that was published after the Chou review is Briggs, which is a case series, N=62. Case series are a type of evidence with high susceptibility to bias.
	27	The NASS 2007 clinical guidelines[16] stated that a study by Riew et al.[17] provided "level II treatment evidence that transforaminal ESI can decrease the likelihood that a patient with radicular leg pain and spinal stenosis will undergo an operation." They also stated that there was grade B recommendation (fair evidence for recommending the intervention) for the statement that "a single radiographically-guided transforaminal epidural steroid injection can produce short-term relief in patients with radiculopathy from lumbar spinal stenosis. ISIS appreciates the opportunity to comment.	Both the guideline and the trial cited were published before the date of the Chou review. See comment #24.

DRAFT

State of Oregon Evidence-based Clinical Guidelines Project

Percutaneous Interventions for Low Back Pain

A clinical practice guideline based on the 2009 American Pain Society Guideline (Interventional Therapies, Surgery, and Interdisciplinary Rehabilitation for Low Back Pain)

June 2012

Oregon Health Authority

Guideline Development Group

Jeanene Smith, MD, MPH; Darren Coffman, MS
Office for Oregon Health Policy and Research

Cat Livingston, MD, MPH
Office for Oregon Health Policy and Research [Clinical Services Consultant]
Oregon Health & Science University

Walter Shaffer, MD
Oregon Division of Medical Assistance Programs

Mylia Christenson
Oregon Health Care Quality Corporation

Steven D. Marks, MD, MHA
PacificSource Health Plans
Oregon Health Leadership Council

Alison Little, MD, MPH; Valerie King, MD, MPH; Catherine Pettinari, PhD; Aasta Thielke, MPH; Mellisa Pensa, MD, MPH; Shannon Vandegriff, BA; Cathy Gordon, MPH
Center for Evidence-based Policy, Oregon Health & Science University

Suggested Citation

Livingston, C., Little, A., King, V., Pettinari, C., Thielke, A., Pensa, M., Vandegriff, S., & Gordon, C. (2012). *State of Oregon Evidence-based Clinical Guidelines Project. Percutaneous interventions for low back pain: A clinical practice guideline based on the 2009 American Pain Society Guideline (Interventional Therapies, Surgery, and Interdisciplinary Rehabilitation for Low Back Pain).* Salem: Office for Oregon Health Policy and Research. Available at: http://www.oregon.gov/OHA/OHPR/HERC/Evidence-Based-Guidelines.shtml

This document was prepared by the Center for Evidence-based Policy at Oregon Health & Science University (the Center) on behalf of the Guideline Development Group and the Office for Oregon Health Policy & Research. This document is intended to help providers, consumers and purchasers of health care in Oregon make informed decisions about health care services. The document is intended as a reference and is provided with the understanding that neither the Center nor the Guideline Development Group are engaged in rendering any clinical, legal, business or other professional advice.

These guidelines should not be construed as dictating an exclusive course of treatment or procedure. Variations in practice may be warranted based on the needs of the individual patient, resources, and limitations unique to the institution or type of practice.

The statements in this document do not represent official policy positions of the Center, the Guideline Development Group, or the Office for Oregon Health Policy and Research. Researchers and authors involved in preparing this document have no affiliations or financial involvement that conflict with material presented in this document.

Objective
This guideline was developed by a collaborative group of public and private partners to provide up-to-date evidence-based guidance on the role of percutaneous interventions in low back pain. The aim of the guideline is to identify evidence-based, appropriate indications for the use of percutaneous interventions in patients with low back pain of any duration, with and without leg pain. This guideline can then be used to create practice standards and coverage guidelines for use across public and private payers. It does not address patients with back pain associated with major trauma, tumor, metabolic disease, inflammatory back disease, fracture, dislocation, major instability or deformity, progressive or severe neurologic deficits, or back pain in children, adolescents or pregnant women. Percutaneous interventions addressed in this guideline include intradiscal, facet joint, sacroiliac joint and epidural steroid injections, prolotherapy, botulinum toxin injections, local injections, medial branch block, radiofrequency denervation, intradiscal electrothermal therapy, percutaneous intradiscal radiofrequency thermocoagulation and coblation nucleoplasty.

Additional evidence concerning other elements of evaluation as well as recommendations for management of low back pain can be found in the State of Oregon Evidence-based Clinical Guidelines:

- Evaluation and Management of Low Back Pain[1]
- Advanced Imaging for Low Back Pain[2]

Background
In June 2009, the Oregon legislature passed health reform legislation HB 2009, which created the Oregon Health Policy Board and charged it with creating a comprehensive health reform plan for our state. In December 2010, the Board released *Oregon's Action Plan for Health*, which lays out "strategies that reflect the urgency of the health care crisis and a timeline for actions that will lead Oregon to a more affordable, world-class health care system." They outlined eight foundational strategies, one of which is to "set standards for safe and effective care." To accomplish this, the plan directs the state to "Identify and develop 10 sets of Oregon-based best practice guidelines and standards that can be uniformly applied across public and private health care to drive down costs and reduce unnecessary care." This work is being conducted by the Oregon Health Services Commission and the Oregon Health Resources Commission in close collaboration with providers, the Center for Evidence-Based Policy, and other key stakeholders.[3]

Development of this guideline:
This guideline was developed by a Guideline Development Group (GDG) consisting of representatives from the State of Oregon Health Authority, the Oregon Healthcare Leadership Council, and the Oregon Corporation for Healthcare Quality with support from clinical evidence specialists from the Center for

---

[1] Livingston, C., King, V., Little, A., Pettinari, C., Thielke, A., & Gordon, C. (2011). *State of Oregon Evidence-based Clinical Guidelines Project. Evaluation and management of low back pain: A clinical practice guideline based on the joint practice guideline of the American College of Physicians and the American Pain Society (Diagnosis and treatment of low back pain).* Salem: Office for Oregon Health Policy and Research.
[2] Livingston, C., Little, A., King, V., Pettinari, C., Thielke, A., Vandegriff, S., & Gordon, C. (2012). *State of Oregon Evidence-based Clinical Guidelines Project. Advanced imaging for low back pain: A clinical practice guideline based on the joint practice guideline of the American College of Physicians and the American Pain Society (Diagnosis and treatment of low back pain).* Salem: Office for Oregon Health Policy & Research.
[3] Effective January 1, 2012, House Bill 2100 (2011) terminates the Health Services Commission and Health Resources Commission and transfers their duties related to evidence-based guideline development to a new Health Evidence Review Commission.

Evidence-based Policy. The Center provided expertise in the process of guideline development and undertook analysis and appraisal to support the development of this guideline.

Methods:
The GDG developed this guideline using the ADAPTÉ[4] framework which is a systematic approach to the endorsement or modification of guideline(s) produced in one cultural context or organizational setting for application in another context. Guideline adaptation is used as an alternative to wholly new guideline development, which can be time consuming, expensive and an inefficient use of resources, when existing quality guidelines are available.

The process for developing this guideline began by searching 17 different databases and other sources for guidelines related to percutaneous interventions for chronic back pain (see appendix A). Candidate guidelines were required to satisfy the following requirements:

- to be evidence-based, that is, guideline recommendations are based on systematic reviews of the literature,
- to address the use of percutaneous interventions in adults with chronic back pain,
- to be published in English and,
- to be freely available to the public.

The GDG required that evidence-based recommendations be made on the basis of both the quality and strength of the underlying evidence from any included guideline's systematic reviews. The initial search identified 10 candidate guidelines which met the above stated criteria (Appendix B). Of the original candidate guidelines, three had been rated as poor quality during the development of a previous guideline and one was excluded because it was not publically available. The six remaining guidelines were then assessed for methodologic quality using a modified AGREE (Appraisal of Guidelines Research and Evaluation) II[5] instrument (Appendix C) by two different guideline quality assessors from the Center for Evidence-based Policy. Two of those guidelines were rated good quality, and one was rated fair with good rigor of development of the evidence and recommendations according to the modified AGREE rating tool. These three guidelines were then examined further for scope and clarity of presentation.

Comparison of the APS guideline was made to the other high quality, comprehensive guidelines, which were produced by the National Institute for Health and Clinical Excellence (NICE), and Towards Optimized Practice, Alberta Clinical Guidelines Program. Of the guidelines considered for review, the GDG felt that the APS guideline was the most comprehensive.

After considering guideline scope and specific modalities addressed, the GDG selected the American Pain Society's 2009 guideline "Interventional therapies, surgery, and interdisciplinary rehabilitation for low back pain: An evidence-based clinical practice guideline from the American Pain Society" as the base guideline, primarily because it had recommendations concerning a broader range of interventions than guidelines from the National Institute for Health and Clinical Evidence (NICE) or from Towards Optimized Practice (TOP). (See Appendix E for procedures addressed in the APS guideline.)

[4] http://www.adapte.org/www/
[5] http://www.agreecollaboration.org/

The APS guideline in its entirety can be found at the following link: http://journals.lww.com/spinejournal/Abstract/2009/05010/Interventional_Therapies,_Surgery,_and.14.aspx. The APS guideline is accompanied by a full systematic review on nonsurgical interventional therapies for low back pain in the same journal issue at: http://www.ampainsoc.org/library/pdf/LBPEvidRev.pdf.

The APS guideline panel arrived at treatment recommendations by first evaluating the evidence for treatments according to a system adapted from the US Preventive Services Task Force for grading the evidence, then estimating the magnitude of effects, including whether the benefits of the treatment outweigh the harms. (See Appendix D for the APS criteria for arriving at recommendations.)

Updating:
The APS guideline was published in 2009. The authors of the guideline were contacted in March 2011 and stated that there had been no new published evidence which would change the recommendations of the guideline and that it was considered current. The GDG recommends that this guideline be reevaluated if the APS issues an updated guideline and at least every two years for currency if the original guideline is not updated.

Recommendations

Below are the recommendations of the APS clinical practice guideline followed by discussion of each recommendation.

Table A. State of Oregon Evidence-based Clinical Guideline Recommendations for Percutaneous Injections of the Spine

Condition	Intervention	Net Benefit	Recommendation	Strength of Recommendation and Quality of Evidence Rating*
Non-radicular Low Back Pain				
Non-specific Low Back Pain	• Prolotherapy	No net benefit	In patients with persistent nonradicular low back pain, **clinicians should not** provide prolotherapy.	Recommendation: Strong Grade: High-quality evidence
	• Local injections • Botulinum toxin injection • Epidural steroid injection • Therapeutic medial branch block • Radiofrequency denervation • Sacroiliac joint steroid injection • Coblation nucleoplasty	Unknown	In patients with persistent nonradicular low back pain, there is insufficient evidence to adequately evaluate the benefits of local injections, botulinum toxic injection, epidural steroid injection, therapeutic medial branch block, radiofrequency denervation, sacroiliac joint steroid injection, or coblation nucleoplasty.	Insufficient evidence to determine net benefits or harms

Condition	Intervention	Net Benefit	Recommendation	Strength of Recommendation and Quality of Evidence Rating*
Presumed discogenic pain	• Intradiscal steroid injection	No net benefit	In patients with presumed discogenic pain, **clinicians should not** provide intradiscal steroid injection.	Recommendation: Strong Grade: High quality-evidence
	• Percutaneous intradiscal radiofrequency thermocoagulation (PIRFT) • Intradiscal electrothermal therapy (IDET)	Unknown	In patients with presumed discogenic pain, there is insufficient evidence to adequately evaluate the benefits of PIRFT or IDET	Insufficient evidence to determine net benefits or harms
Presumed facet joint pain	• Facet joint steroid injection	No net benefit	In patients with presumed facet joint pain, **clinicians should not** provide facet joint steroid injection.	Recommendation: Strong Grade: Moderate-quality evidence
	• Radiofrequency denervation	Unknown	In patients with presumed facet joint pain, there is insufficient evidence to adequately evaluate the benefits of radiofrequency denervation.	Insufficient evidence to determine net benefits or harms
Presumed sacroiliac joint pain	• Sacroiliac joint steroid injection	Unknown	In patients with presumed sacroiliac joint pain, there is insufficient evidence to adequately evaluate the benefits of sacroiliac joint steroid injection.	Insufficient evidence to determine net benefits or harms
Radiculopathy or Spinal Stenosis				
Radiculopathy with herniated lumbar disc	• Epidural steroid injection	**Moderate benefit (short-term)**	In patients with persistent radiculopathy due to herniated lumbar disc, **clinicians should** discuss the risks and benefits of epidural steroid injections as an option. It is recommended that Shared decision-making regarding epidural steroid injection includes a specific discussion about inconsistent evidence showing moderate short-term benefits and lack of long-term benefits.	Recommendation: Weak Grade: Moderate-quality evidence

Condition	Intervention	Net Benefit	Recommendation	Strength of Recommendation and Quality of Evidence Rating*
Radiculopathy with herniated lumbar disc, cont.	• Coblation nucleoplasty	Unknown	In patients with radiculopathy with herniated lumbar disc, there is insufficient evidence to adequately evaluate the benefits.	Insufficient evidence to determine net benefits or harms
Radiculopathy	• Radiofrequency denervation	Unknown	In patients with radiculopathy, there is insufficient evidence to adequately evaluate the benefits.	Insufficient evidence to determine net benefits or harms
Symptomatic Spinal Stenosis	• Epidural steroid injection	Unknown	In patients with spinal stenosis, there is insufficient evidence to adequately evaluate the benefits.	Insufficient evidence to determine net benefits or harms

***See Appendix D for complete description of APS and ACP evidence grading methods. Chou, et al. (2009) utilize the US Prevent Services Task Force criteria for rating the strength of recommendation and quality of evidence. Recommendations in this table are modified to fit GRADE terminology for consistency among State of Oregon guidelines.**

Recommendation #1[6]:

Epidural Steroid Injection for persistent radiculopathy due to herniated lumbar disc

> *In patients with persistent radiculopathy due to herniated lumbar disc, it is recommended that clinicians discuss risks and benefits of epidural steroid injection as an option (weak recommendation, moderate-quality evidence). It is recommended that shared decision-making regarding epidural steroid injection include a specific discussion about inconsistent evidence showing moderate short-term benefits, and lack of long-term benefits. There is insufficient evidence to adequately evaluate benefits and harms of epidural steroid injection for spinal stenosis.*

For radiculopathy due to herniated lumbar disc, evidence on benefits of epidural steroid injection is mixed. Although some higher-quality trials (Arden 2005; Bush 1991; Dilke 1973; Wilson-MacDonald 2005) found epidural steroid injection associated with moderate short-term (through up to 6 weeks) benefits in pain or function, others (Carette 1997; Karppinen 2001; Ng 2005) found no differences *versus* placebo injection. Reasons for the discrepancies between trials is uncertain, but could be related to the type of comparator treatment, as trials (Beliveau 1971; Breivik 1976; Bush 1991; Carette 1997; Cuckler 1985; Karppinen 2001; Klenerman 1984; Ng 2005; Rogers 1992; Snoek 1977; Zahaar 1991) that compared an epidural steroid injection to an epidural saline or local anesthetic injection tended to report poorer results than trials (Arden 2005; Dilke 1973; Helliwell 1985; Mathews 1987; Ridley 1988; Wilson-MacDonald 2005) that compared an epidural steroid injection to a soft-tissue (usually interspinous ligament) placebo injection. Regardless of the comparator intervention, there is no

[6] Extracted and modified from Chou, et. al. (2009)

convincing evidence that epidural steroids are associated with long-term benefits and most trials (Arden 2005; Carette 1997; Riew 2000; Wilson-MacDonald 2005) found no reduction in rates of subsequent surgery. Although serious complications following epidural steroid injection are rare in clinical trials, (Arden 2005; Karppinen 2001; Kolsi 2000; Kraemer 1997; Ng 2005) there are case reports of paralysis and infections. (Glaser 2005; Hooten 2006; Huntoon 2004) There is insufficient evidence on clinical outcomes to recommend a specific approach for performing epidural steroid injection (Ackerman 2007; Kolsi 2000; Kraemer 1997; McGregor 2001; Thomas 2003) or on use of fluoroscopic guidance. In addition, insufficient evidence exists to recommend how many epidural injections to perform, though 1 higher-quality trial found that if an initial epidural steroid injection did not result in benefits, additional injections over a 6-week period did not improve outcomes (Arden 2005).

Epidural steroid injection for the treatment of radiculopathy with herniated lumbar disc is the only percutaneous intervention found to have a net benefit, and the benefit appears to be short-term.

Decisions regarding use of epidural steroid injection should be based on a shared decision-making process that includes a discussion of the inconsistent evidence for short-term benefit, lack of long-term benefit, potential risks, and costs. Patient preferences and individual factors should also be considered. For example, epidural steroid injection may be a reasonable option for short-term pain relief in patients who are less optimal surgery candidates due to comorbidities. There is insufficient evidence to guide specific recommendations for timing of epidural steroid injection, though most trials enrolled patients with at least subacute (greater than 4 weeks) symptoms.

Evidence on efficacy of epidural steroid injection for spinal stenosis is sparse and shows no clear benefit, though more trials are needed to clarify effects (Cuckler 1985; Fukusaki 1998; Zahaar 1991). Although chymopapain chemonucleolysis (see glossary, Supplemental Digital Content 1, http://links.lww.com/A840) is effective for radiculopathy due to herniated lumbar disc, (Gibson 2007a, 2007b) it is less effective than discectomy (see glossary, Supplemental Digital Content 1, http://links.lww.com/A840) and is no longer widely available in the United States, in part due to risk of severe allergic reactions.

Recommendation #2[7]:
Facet Joint Injection, Prolotherapy, Intradiscal Corticosteroid Injection

> *In patients with persistent nonradicular low back pain, facet joint corticosteroid injection, prolotherapy, and intradiscal corticosteroid injection* **are not recommended** *(strong recommendation, moderate-quality evidence).*

Injections and most interventional therapies for nonradicular low back pain target specific areas of the back that are potential sources of pain, including the muscles and soft tissues (botulinum toxin injection, prolotherapy, and local injections [see glossary, Supplemental Digital Content 1, http://links.lww.com/A840]), facet joints (facet joint steroid injection, therapeutic medial branch block, and radiofrequency denervation [see glossary, Supplemental Digital Content 1, http://links.lww.com/A840]), degenerated intervertebral discs (intradiscal steroid injection, IDET, [see glossary, Supplemental Digital Content 1, http://links.lww.com/A840] and related procedures), and sacroiliac joints (sacroiliac joint injection)

[7] Extracted and modified from Chou, et. al. (2009)

There is no convincing evidence from randomized trials that injections and other interventional therapies are effective for nonradicular low back pain. Facet joint steroid injection (Carette 1991; Lilius 1989) prolotherapy (Dagenais 2007) and intradiscal steroid injections (Khot 2004; Simmons 1992) are not recommended because randomized trials consistently found them to be no more effective than sham therapies.

Five randomized, placebo-controlled trials evaluated prolotherapy (Gibson 2007a; Huntoon 2004; Klenerman 1984; Malmivaara 2007; Weber 1983). All were included in a higher quality Cochrane review (Willems 2004). Four trials were rated higher quality (Huntoon 2004; Klenerman 1984; Malmivaara 2007; Weber 1983). For chronic nonspecific low back pain, 3 trials (2 higher quality: Klenerman 1984, Malmivaara 2007) found no difference between prolotherapy and either saline or local anesthetic control injections for short-or long-term (up to 24 months) pain or disability (Malmivaara 2007).

Recommendation #3[8]:
Other Interventional Procedures

> *There is insufficient evidence to adequately evaluate benefits of local injections, botulinum toxin injection, epidural steroid injection, intradiscal electrothermal therapy (IDET), therapeutic medial branch block, radiofrequency denervation, sacroiliac joint steroid injection, coblation nucleoplasty, percutaneous intradiscal radiofrequency thermocoagulation …. or other medications for nonradicular low back pain.*

For local injections, there is insufficient evidence to accurately judge benefits because available trials are small, lower-quality, and evaluate heterogeneous populations and interventions (Collee 1991; Garvey 1989; Hameroff 1981; Sonne 1985). Trials of IDET (Freeman 2005; Pauza 2004) and radiofrequency denervation (Leclaire 2001; Nath 2008; van Kleef 1999; van Wijk 2005) reported inconsistent results. There were a small number of higher quality trials, and in the case of radiofrequency denervation, the trials had technical or methodologic shortcomings (Hooten 2005), making it difficult to reach conclusions about benefits**.** For other interventional therapies, data are limited to 1-2 small placebo-controlled randomized trials (botulinum toxin injection (Foster 2001), epidural steroid injection for nonradicular low back pain (Serrao 1992), PIRFT (Barendse 2001, Ercelen 2003) and sacroiliac joint steroid injection [see glossary, Supplemental Digital Content 1, http://links.lww.com/A840] (Luukkainen 2002), or there are no placebo-controlled randomized trials (therapeutic medial branch block, coblation nucleoplasty….or other medications).

[8] Extracted and modified from Chou, et. al. (2009)

Appendix A. Sources Searched for Low Back Pain Guidelines

1. British Medical Journal – Clinical Evidence
2. Cochrane Library
3. Agency for Healthcare Research and Quality
4. ECRI
5. Hayes, Inc
6. Veterans Administration – Technology Assessment Program (VA TAP)
7. Blue Cross Blue Shield HTA
8. Centers for Medicare and Medicaid
9. CADTH
10. Washington HTA Program
11. US Preventive Services Task Force
12. ICSI
13. Guidelines.gov
14. American College of Physicians AND American Pain Society
15. American Physical Therapy Association
16. PEDro.org.au (evidence-based physiotherapy database)
17. GIN Guidelines Database

Appendix B. Low Back Pain Guidelines Identified

Methods Summary:
Initially, 17 databases and other sources for guidelines related to percutaneous Interventions for low back pain were searched. Candidate guidelines were required to:

- be evidence-based (recommendations based on a full systematic review)
- be comprehensive
- be published in English
- be freely available to the public

Ten candidate guidelines were identified, of which six were sufficiently comprehensive and were assessed by two clinical epidemiologists for methodologic quality using a modified AGREE (Appraisal of Guidelines Research and Evaluation) II[9] instrument.
Candidate guidelines were then assessed considering:

- age
- source
- specific treatment elements addressed
- presentation

The GDG selected the guideline of highest quality and that was most comprehensive. (See guideline text for comprehensive Methods discussion)

Low Back Pain Guidelines Identified in Search – Selected for Quality Assessment

Armon, C., Argoff, C.E., Samuels, J., Backonja, M.M. (2007). Assessment: Use of epidural steroid injections to treat radicular lumbosacral pain: Report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. Neurology 68:723-729.
Overall guideline quality rating: Fair

Chou, R., Loesser, J.D., Owens, D.K., Rosenquist, R.W., Atlas, S.J., Baisden, J., Carragee, E.J., Grabois, M., Murphy, D.R., Resnick, D.K., Stanos, S.P., Shaffer, W.O., Wall E.M. (2009) Interventional therapies, surgery, and interdisciplinary rehabilitation for low back pain: An evidence-based clinical practice guideline from the American Pain Society. Spine 34:10:1066-1077. – accompanied by:
Chou, R., Atlas, S.J., Stanos, S.P., Rosenquist, R.W. (2009). A review of the evidence for an American Pain Society clinical practice guideline. Spine 34:10:1078-1094.
Overall guideline quality rating: Fair with good rigor of development of evidence and recommendations

Manchikanti, L ., Boswell, M.V., Singh, V., Benyamin, R.M., Fellows, B., Abdi, S., Buenaventura, R.M., Conn, A., Datta, S., Derby, R., Falco, F.J.E., Erhart, S., Diwan, S., Hayek, S.M., Helm II, S., Parr, A.T., Schultz, D.M., Smith, H.S., Wolfer, L. R., Hirsch, J.A. (2009). Comprehensive evidence-based guidelines for interventional techniques in the management of chronic spinal pain. Pain Physician 12:699-802.
Overall guideline quality rating: Poor

National Health and Medical Research Council. Australian Acute Musculoskeletal Pain Guidelines Group. (2003). Evidence-based management of acute musculoskeletal pain. (Website states that status is "current"). [Chapter 4 of document is on Acute Low Back Pain.]
http://www.nhmrc.gov.au/_files_nhmrc/file/publications/synopses /cp94.pdf
Overall guideline quality rating: Fair

[9] http://www.agreecollaboration.org/

National Institute for Health and Clinical Excellence (NICE). (2009). Low back pain: Early management of persistent non-specific low back pain. London, UK: National Institute for Health and Clinical Excellence. Retrieved September 30, 2010, from http://www.nice.org.uk/nicemedia/live/11887/44343/44343.pdf
Overall guideline quality rating: Good

Towards Optimized Practice. (2009). Management of low back pain. Edmonton, AB: Towards Optimized Practice Program.
Overall guideline quality rating: Good

Low Back Pain Guidelines Identified in Search– Not Selected for Quality Assessment

American College of Occupational and Environmental Medicine (ACOEM). (2007). Low back disorders. Occupational medicine practice guidelines: Evaluation and management of common health problems and functional recovery in workers. 2nd ed. Elk Grove Village, IL: ACOEM.
Overall guideline quality rating: Fair

Institute for Clinical Systems Improvement (ICSI). (2010). Adult low back pain. Fourteenth edition. Bloomington, MN: ICSI.
Overall guideline quality rating: Poor

Michigan Quality Improvement Consortium. (2008). Management of acute low back pain. Southfield, MI: Michigan Quality Improvement Consortium.
Overall guideline quality rating: Poor

University of Michigan Health System. (2010). Acute low back pain. Ann Arbor, MI: University of Michigan Health System.
Overall guideline quality rating: Poor

Appendix C: Methodology Checklist Adapted from the AGREE II materials

Methodology Checklist: Guidelines				
Guideline citation *(Include name of organization, title, year of publication, journal title, pages)*				
Guideline Topic:				
Checklist completed by:			Date:	
SECTION 1: PRIMARY CRITERIA				
To what extent is there		***Assessment/Comments:***		
1.1	**RIGOR OF DEVELOPMENT: Evidence** • Systematic literature search • Study selection criteria clearly described • Quality of individual studies and overall strength of the evidence assessed • Explicit link between evidence & recommendations *(If any of the above are missing, rate as poor)*	GOOD	FAIR	POOR
1.2	**RIGOR OF DEVELOPMENT: Recommendations** • Methods for developing recommendations clearly described • Strengths and limitations of evidence clearly described • Benefits/side effects/risks considered • External review	GOOD	FAIR	POOR
1.3	**EDITORIAL INDEPENDENCE**[10] • Views of funding body have not influenced the content of the guideline • Competing interests of members have been recorded and addressed	GOOD	FAIR	POOR
If any of three primary criteria are rated poor, the entire guideline should be rated poor.				
SECTION 2: SECONDARY CRITERIA				
2.1	**SCOPE AND PURPOSE** • Objectives described • Health question(s) specifically described • Population (patients, public, etc.) specified	GOOD	FAIR	POOR

[10] Editorial Independence is a critical domain. However, it is often very poorly reported in guidelines. The assessor should not rate the domain, but write "unable to assess" in the comment section. If the editorial independence is rated as "poor", indicating a high likelihood of bias, the entire guideline should be assessed as poor.

SECTION 2: SECONDARY CRITERIA, Cont.		
2.2	**STAKEHOLDER INVOLVEMENT** • Relevant professional groups represented • Views and preferences of target population sought • Target users defined	GOOD FAIR POOR
2.3	**CLARITY AND PRESENTATION** • Recommendations specific, unambiguous • Management options clearly presented • Key recommendations identifiable • Application tools available • Updating procedure specified	GOOD FAIR POOR
2.4	**APPLICABILITY** • Provides advice and/or tools on how the recommendation(s) can be put into practice • Description of facilitators and barriers to its application • Potential resource implications considered • Monitoring/audit/review criteria presented	GOOD FAIR POOR
SECTION 3: OVERALL ASSESSMENT OF THE GUIDELINE		
3.1	How well done is this guideline?	GOOD FAIR POOR
3.2	Other reviewer comments:	

Description of Ratings: Methodology Checklist for Guidelines

The checklist for rating guidelines is organized to emphasize the use of evidence in developing guidelines and the philosophy that "evidence is global, guidelines are local." This philosophy recognizes the unique situations (e.g., differences in resources, populations) that different organizations may face in developing guidelines for their constituents. The second area of emphasis is transparency. Guideline developers should be clear about how they arrived at a recommendation and to what extent there was potential for bias in their recommendations. For these reasons, rating descriptions are only provided for the primary criteria in section one. There may be variation in how individuals might apply the good, fair, and poor ratings in section two based on their needs, resources, organizations, etc.

Section 1. Primary Criteria (rigor of development and editorial independence) ratings:

Good: All items listed are present, well described, and well executed (e.g., key research references are included for each recommendation).
Fair: All items are present, but may not be well described or well executed.
Poor: One or more items are absent or are poorly conducted

Appendix D. APS Guideline Criteria for Treatment Recommendations

The APS guideline panel arrived at treatment recommendations by first evaluating the evidence for treatments according to a system adapted from the US Preventive Services Task Force for grading the evidence, then estimating the magnitude of effects, including whether the benefits of the treatment outweigh the harms.

The underlying strength of the evidence for each intervention was given a rating of good, fair or poor based on factors such as the quality, quantity, consistency, and generalizability of the evidence (Table 1).

Table 1. APS Criteria for Grading the Strength of Evidence

Rating	Strength
Good	Evidence includes consistent results from well-designed, well-conducted studies in representative populations that directly assess effects on health outcomes (at least 2 consistent, higher-quality trials)
Fair	Evidence is sufficient to determine effects on health outcomes, but the strength of the evidence is limited by the number, quality, size, or consistency of included studies; generalizability to routine practice; or indirect nature of the evidence on health outcomes (at least 1 higher-quality trial of sufficient sample size; 2 or more higher-quality trials with some inconsistency; at least 2 consistent, lower-quality trials, or multiple consistent observational studies with no significant methodologic flaws
Poor	Evidence is insufficient to assess effects on health outcomes because of limited number or power of studies, large and unexplained inconsistency between higher-quality trials, important flaws in trial design or conduct, gaps in the chain of evidence, or lack of information on important health outcomes.

Depending on the strength of the evidence for an intervention, the APS used the following criteria for making a recommendation.

Table 2. APS Criteria for making treatment recommendations

Grade	Criteria for making a recommendation
A	The panel strongly recommends that clinicians consider offering the intervention to eligible patients. The panel found good evidence that the intervention improves health outcomes and concludes that benefits substantially outweigh harms.
B	The panel recommends that clinicians consider offering the intervention to eligible patients. The panel found at least fair evidence that the intervention improves health outcomes and concludes that benefits moderately outweigh harms, or that benefits are small but there are no significant harms, costs, or burdens associated with the intervention.
C	The panel makes no recommendation for or against the intervention. The panel found at least fair evidence that the intervention can improve health outcomes, but concludes that benefits only slightly outweigh harms, or the balance of benefits and harms is too close to justify a general recommendation.
D	The panel recommends against offering the intervention. The panel found at least fair evidence that the intervention is ineffective or that harms outweighs benefits.
I	The panel found insufficient evidence to recommend for or against the intervention. Evidence that the intervention is effective is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.

If a recommendation was made, the APS assigned an overall grade of its strength, adapting the grading system of the international Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) working group. Strong recommendations are required to have clear evidence of benefit or harm. Weak recommendations are based on finely balanced benefits, risks and burdens.

Table 3. ACP Clinical Practice Guidelines Grading System[11]

Quality of Evidence	Strength of Recommendation	
	Benefits Do or Do Not Clearly Outweigh Risks	**Benefits and Risks and Burdens Are Finely Balanced**
High	Strong	Weak
Moderate	Strong	Weak
Low	Strong	Weak
Insufficient evidence to determine net benefits or harms		

The ACP/APS guideline panel considered interventions to have "proven" benefit if there was at least fair quality evidence of moderate or substantial benefit (or of small benefit with no significant harms, costs or burdens).

[11] Adapted from the system developed by the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) workshop by the American College of Physicians.

Appendix E. Treatments addressed in APS guideline*

Treatment	Definitions Procedures are defined according to APS http://links.lww.com/A840
Prolotherapy (sclerotheraphy) Injections	A procedure involving the repeated injection of an irritant chemical into the soft tissues of the back in order to provoke an inflammatory response that will theoretically subsequently lead to strengthening of the soft tissues with decrease in pain and disability. Also referred to as sclerotherapy
Facet joint corticosteroid injections	Injection of corticosteroid into the facet joints.
Therapeutic medial branch block	Injection of local anesthetic with or without corticosteroid in the area of the medial branch of the posterior primary ramus, the primary nerve innervating the intervertebral facet joint. Usually used as a diagnostic procedure to identify facet joint pain, but has also been used as a therapeutic procedure
Intradiscal corticosteroid injections	Injection of corticosteroid into the intervertebral disc.
Radiofrequency denervation	A procedure involving the destruction of nerves using heat generated by a radiofrequency current.
Intradiscal electrothermal therapy (IDET)	A procedure involving the placement of an electrode or catheter into the intervertebral disc annulus or nucleus and applying electrothermal energy to alter adjacent pain receptors or other structures.
Epidural steroid injection	Injection of corticosteroids via a catheter into the space between the dura and the spine. Common approaches for administering epidural steroid injections are through the interlaminar space, via the neuroforamen under fluoroscopic guidance (transforaminal), and through the sacral hiatus at the sacral canal (caudal).
Local injections	Injection of local anesthetic (with or without corticosteroid) into the muscles or soft tissues of the back. Trigger point injections, a type of local injection, involve an injection performed at a tender area, often with a palpable nodule or band.
Sacroiliac joint steroid Injection	Injection of corticosteroid into or around the sacroiliac joint.
Botulinium toxin injection	Injection of botulinum toxin (an antispasmodic) into the muscles of the back.
Chemonucleolysis	Treatment of herniated discs with intradiscal injections of a proteolysis enzyme, most commonly chymopapain (an extract from papaya). Chymopapain acts by digesting the jelly-like inner portion of the disc known as the nucleus pulposus, while at the same time, leaving the outer portion, the annulus fibrosis, essentially intact.
Adhesiolysis and forceful epidural injection	(not defined)
Coblation® nucleoplasty	A procedure involving the use of a bipolar radiofrequency current in order to create a series of channels in an intervertebral disc and reduce the volume of tissue.
Percutaneous intradiscal radiofrequency thermocoagulation (PIRFT)	A procedure involving the placement of an electrode of catheter into the intervertebral disc and applying alternating radiofrequency current. Sometimes classified as a variant of intradiscal electrothermal therapy (IDET).

*Chou, R., Loesser, J.D., Owens, D.K., et al. (2009). Interventional therapies, surgery, and interdisciplinary rehabilitation for low back pain: An evidence-based clinical practice guideline from the American Pain Society. *Spine*, 34(10):1066-1077.

Appendix F. List of Peer Reviewers

Invited: Accepted & Reviewed

Susan Bamberger, PT, MPT, DIP MDT
Past President
Oregon Physical Therapy Association

Roger Chou, MD
Scientific Director
Oregon Evidence-based Practice Center
Division of General Internal Medicine and Geriatrics
Oregon Health & Science University

Timothy J. Craven, MD, MPH
Associate Medical Director
Providence Health Plan MCO

Rick Deyo, MD, MPH
Kaiser Permanente Professor of Evidence-Based Family Medicine
Director, KL2 Multidisciplinary Clinical Research Career Development Program
Director, OCTRI Community and Practice-based Research Program
Departments of Family Medicine and Internal Medicine
Oregon Health & Science University

Marc Gosselin, MD
Associate Professor
Director, Thoracic Imaging
Department of Diagnostic Radiology
Oregon Health & Science University

Luci Kovacevic, MD, MPH
Occupational Medicine Physician
Cascade Medical Associates

David Pass, MD
Anesthesiologist
Medical Director
Providence Health Plans

LaVerne A. Saboe, Jr., DC, DACAN, FICC, DABFP, FACO
Chiropractic Physician
Past president, Chiropractic Association of Oregon

Invited: Declined/Did Not Respond/Did Not Review

Fourteen additional reviewers were invited but either declined, did not respond, missed the deadline or did not return the review. Areas of professional expertise for invited reviewers included:

Anesthesiology
Behavioral Health
Complementary and Alternative Medicine
Family Medicine
Internal Medicine
Occupational Medicine
Orthopedic Surgery
Neurosurgery
Pain Advocacy
Pain Medicine
Physical Therapy
Physical Medicine and Rehabilitation
Radiology
Sports Medicine
Worker's Compensation

Appendix G. References

Ackerman WE III, Ahmad M. (2007). The efficacy of lumbar epidural steroid injections in patients with lumbar disc herniations. *Anesth Analg,* 104:1217–22.

Arden NK, Price C, Reading I, et al. (2005). A multicentre randomized controlled trial of epidural corticosteroid injections for sciatica. *Rheumatology,* 44:1399–406.

Barendse GA, van Den Berg SG, Kessels AH, et al. (2001). Randomized controlled trial of percutaneous intradiscal radiofrequency thermocoagulation for chronic discogenic back pain. *Spine*, 26:287–92.

Beliveau P. (1971). A comparison between epidural anaesthesia with and without corticosteroid in the treatment of sciatica. *Rheumatol Phys Med,* 11:40–3.

Breivik H, Hesla PE, Molnar I, et al. (1976). Treatment of chronic low back pain and sciatica: comparison of caudal epidural injections of bupivicaine and methylprednisolone with bupivicaine followed by saline. *Adv Pain Res Ther,* 1:927–32.

Bush K, Hillier S. (1991). A controlled study of caudal epidural injections of triamcinolone plus procaine for the management of intractable sciatica. *Spine,* 16:572–5.

Carette S, Marcoux S, Truchon R, et al. (1991). A controlled trial of corticosteroid injections into facet joints for chronic low back pain. *N Engl J Med,* 325:1002–7.

Carette S, Leclaire R, Marcoux S, et al. (1997). Epidural corticosteroid injections for sciatica due to herniated nucleus pulposus. *N Engl J Med,* 336:1634–40.

Chou, R., Loesser, J.D., Owens, D.K., et al. (2009). Interventional therapies, surgery, and interdisciplinary rehabilitation for low back pain: An evidence-based clinical practice guideline from the American Pain Society. *Spine*, 34(10):1066-1077.

Collee G, Dijkmans B, Vandenbroucke J, et al. (1991). Iliac crest pain syndrome in low back pain. A double blind, randomized study of local injection therapy. *J Rheumatol,* 18:1060–63.

Cuckler J, Bernini P, Wiesel S, et al. (1985). The use of epidural steroids in the treatment of lumbar radicular pain. *J Bone Joint Surg,* 67A:63–6.

Dagenais S, Yelland MJ, Del Mar C, et al. (2007). Prolotherapy injections for chronic low-back pain. *Cochrane Database Syst Rev,* CD004059.

Dilke TFW, Burry HC, Grahame R. (1973). Extradural corticosteroid injection in management of lumbar nerve root compression. *BMJ,* 16:635–7.

Ercelen O, Bulutcu E, Oktenoglu T, et al. (2003). Radiofrequency lesioning using two different time modalities for the treatment of lumbar discogenic pain: a randomized trial. *Spine,* 28:1922–7.

Foster L, Clapp L, Erickson M, et al. (2001). Botulinum toxin A and chronic low back pain: a randomized, double-blind study. *Neurology,* 56:1290–93.

Freeman BJ, Fraser RD, Cain CM, et al. (2005). A randomized, double-blind, controlled trial: intradiscal electrothermal therapy versus placebo for the treatment of chronic discogenic low back pain. *Spine,* 30:2369–77.

Fukusaki M, Kobayashi I, Hara T, et al. (1998). Symptoms of spinal stenosis do not improve after epidural steroid injection. *Clin J Pain,* 14:148–51.

Garvey T, Marks M, Wiesel S. (1989). A prospective, randomized, double-blind evaluation of trigger-point injection therapy for low-back pain. *Spine,* 14:962–4.

Gibson JN. (2007a). Surgical interventions for lumbar disc prolapse: updated Cochrane Review. *Spine,* 32:1735–47.

Gibson J, Waddell G. (2007b). Surgical interventions for lumbar disc prolapse. *Cochrane Database Syst Rev,* CD001350.

Glaser SE, Falco F. (2005). Paraplegia following a thoracolumbar transforaminal epidural steroid injection. *Pain Physician,* 8:309–14.

Hameroff S, Crago B, Blitt C, et al. (1981). Comparison of bupivacaine, etidocaine, and saline for trigger-point therapy. *Anaesth Analg,* 60:752–5.

Helliwell M. (1985). Outpatient treatment of low back pain and sciatica by a single extradural corticosteroid injection. *Br J Clin Pract,* 39:228–31.

Hooten WM, Martin DP, Huntoon MA. (2005). Radiofrequency neurotomy for low back pain: evidence-based procedural guidelines. *Pain Med,* 6:129–38.

Hooten WM, Mizerak A, Carns PE, et al. (2006). Discitis after lumbar epidural corticosteroid injection: a case report and analysis of the case report literature. *Pain Med,* 7:46–51.

Huntoon MA, Martin DP. (2004). Paralysis after transoforaminal epidural injection and previous spinal surgery. *Reg Anesth Pain Med,* 29:494–5.

Karppinen J, Malmivaara A, Kurunlahti M, et al. (2001). Periradicular infiltration for sciatica: a randomized controlled trial. *Spine,* 26:1059–67.

Khot A, Bowditch M, Powell J, et al. (2004). The use of intradiscal steroid therapy for lumbar spinal discogenic pain: a randomized controlled trial. *Spine,* 29:833–6.

Kolsi I, Delecrin J, Berthelot JM, et al. (2000). Efficacy of nerve root versus interspinous injections of glucocorticoids in the treatment of disk-related sciatica. A pilot, prospective, randomized, double-blind study. *Joint Bone Spine,* 67:113–18.

Klenerman L, Greenwood R, Davenport H, et al. (1984). Lumbar epidural injections in the treatment of sciatica. *Br J Rheum,* 23:35–8.

Kraemer J, Ludwig J, Bickert U, et al. (1997). Lumbar epidural perineural injection: a new technique. *Eur Spine J,* 6:357–61.

Leclaire R, Fortin L, Lambert R, et al. (2001). Radiofrequency facet joint denervation in the treatment of low back pain: a placebo-controlled clinical trial to assess efficacy. *Spine,* 26:1411–16.

Lilius G, Lassonen AM, Myllynen P, et al. (1989). The lumbar facet joint syndrome–significance of inappropriate signs. A randomized, placebocontrolled trial. *French J Orthop Surg,* 3:479–86.

Luukkainen R, Wennerstrand P, Kautiainen H, et al. (2002). Efficacy of periarticular corticosteroid treatment of the sacroiliac join in nonspondyloarthropathic patients with chronic low back pain in the region of the sacroiliac joint. *Clinical Exp Rheumatol,* 20:52–4.

Malmivaara A, Slatis P, Heliovaara M, et al. (2007). Surgical or nonoperative treatment for lumbar spinal stenosis? A randomized controlled trial. *Spine,* 32:1–8.

Mathews J, Mills S, Jenkins V, et al. (1987). Back pain and sciatica: controlled trials of manipulation, traction, sclerosant and epidural injection. *Br J Rheumatol*, 26:416–23.

McGregor AH, Anjarwalla NK, Stambach T. (2001). Does the method of injection alter the outcome of epidural injections? *J Spinal Disord,* 14:507–10.

Nath S, Nath C, Pettersson K. (2008). Percutaneous lumbar zygapophysial (facet) joint neurotomy using radiofrequency current, in the management of chronic low back pain. *Spine,* 33:1291–7.

Ng L, Chaudhary N, Sell P. (2005). The efficacy of corticosteroids in periradicular infiltration for chronic radicular pain. A randomized, double-blind, controlled trial. *Spine,* 30:857–62.

Pauza KJ, Howell S, Dreyfuss P, et al. (2004). A randomized, placebo-controlled trial of intradiscal electrothermal therapy for the treatment of discogenic low back pain. *Spine J,* 4:27–35.

Ridley M, Kingsley G, Gibson T, et al. (1988). Outpatient lumbar epidural corticosteroids in the management of sciatica. *Br J Rheumatol,* 27:295–9.

Riew K, Yin Y, Gilula L, et al. (2000). The effect of nerve-root injections on the need for operative treatment of lumbar radicular pain. A prospective, randomized, controlled, double-blind study. *J Bone Joint Surg,* 82-A:1589– 93.

Rogers P, Nash T, Schiller D, et al. (1992). Epidural steroids for sciatica. *Pain Clin*, 5:67–72.

Serrao JM, Marks RL, Morley SJ, et al. (1992). Intrathecal midazolam for the treatment of chronic mechanical low back pain: a controlled comparison with epidural steroid in a pilot study. *Pain,* 48:5–12.

Simmons JW, McMillin JN, Emery SF, et al. (1992). Intradiscal steroids. A prospective double-blind clinical trial. *Spine,* 17(suppl 6):S172–5.

Sonne M, Chirstensen K, Hansen S, et al. (1985). Injection of steroids and local anaesthetics as therapy for low-back pain. *Scand J Rheum,* 14:343–5.

Snoek W, Weber H, Jorgensen B. (1977). Double blind evaluation of extradural methyl prednisolone for herniated lumbar discs. *Acta Orthop Scand,* 48:635–41.

Thomas E, Cyteval C, Abiad L, et al. (2003). Efficacy of transforaminal versus interspinous corticosteroid injections in discal radiculalgia–a prospective, randomized, double-blind study. *Clin Rheumatol,* 22:299–304.

van Kleef M, Barendse G, Kessels A, et al. (1999). Randomized trial of radiofrequency lumbar facet denervation for chronic low back pain. *Spine,* 24:1937–42.

van Wijk R, Geurts J, Wynne H, et al. (2005). Radiofrequency denervation of lumbar facet joints in the treatment of chronic low back pain: a Randomized, Double-Blind, Sham Lesion-Controlled Trial. *Clin J Pain,* 21:335–44.

Weber H. (1983). Lumbar disc herniation. A controlled, prospective study with ten years of observation. *Spine,* 8:131–40.

Willems P, Jacobs W, Duinkerke E, et al. (2004). Lumbar discography: should we use prophylactic antibiotics? A study of 435 consecutive discograms and a systematic review of the literature. *J Spinal Disord Tech,* 17:243–7.

Wilson-MacDonald J, Burt G, Griffin D, et al. (2005). Epidural steroid injection for nerve root compression. A randomised, controlled trial. *J Bone Joint Surg Br,* 87:352–5.

Zahaar M. (1991). The value of caudal epidural steroids in the treatment of lumbar neural compression syndromes. *J Neurol Orthop Med Surg,* 12:181–4.

SPINE Volume 34, Number 10, pp 1078–1093
©2009, Lippincott Williams & Wilkins

Nonsurgical Interventional Therapies for Low Back Pain

A Review of the Evidence for an American Pain Society Clinical Practice Guideline

Roger Chou, MD,*† Steven J. Atlas, MD, MPH,‡ Steven P. Stanos, DO,§ and Richard W. Rosenquist, MD¶

Study Design. Systematic review.

Objective. To systematically assess benefits and harms of nonsurgical interventional therapies for low back and radicular pain.

Summary of Background Data. Although use of certain interventional therapies is common or increasing, there is also uncertainty or controversy about their efficacy.

Methods. Electronic database searches on Ovid MEDLINE and the Cochrane databases were conducted through July 2008 to identify randomized controlled trials and systematic reviews of local injections, botulinum toxin injection, prolotherapy, epidural steroid injection, facet joint injection, therapeutic medial branch block, sacroiliac joint injection, intradiscal steroid injection, chemonucleolysis, radiofrequency denervation, intradiscal electrothermal therapy, percutaneous intradiscal radiofrequency thermocoagulation, Coblation nucleoplasty, and spinal cord stimulation. All relevant studies were methodologically assessed by 2 independent reviewers using criteria developed by the Cochrane Back Review Group (for trials) and by Oxman (for systematic reviews). A qualitative synthesis of results was performed using methods adapted from the US Preventive Services Task Force.

Results. For sciatica or prolapsed lumbar disc with radiculopathy, we found good evidence that chemonucleolysis is moderately superior to placebo injection but inferior to surgery, and fair evidence that epidural steroid injection is moderately effective for short-term (but not long-term) symptom relief. We found fair evidence that spinal cord stimulation is moderately effective for failed back surgery syndrome with persistent radiculopathy, though device-related complications are common. We found good or fair evidence that prolotherapy, facet joint injection, intradiscal steroid injection, and percutaneous intradiscal radiofrequency thermocoagulation are not effective. Insufficient evidence exists to reliably evaluate other interventional therapies.

Conclusion. Few nonsurgical interventional therapies for low back pain have been shown to be effective in randomized, placebo-controlled trials.

Key words: low back pain, systematic review, injections, spinal, electrical stimulation therapy, intervertebral disc, electrocoagulation, intervertebral disc chemolysis, injections, epidural, denervation. **Spine 2009;34:1078–1093**

A variety of nonsurgical interventional procedures are used for treatment of low back pain. These include injections of medications, irritants, or proteolytic enzymes into soft tissues outside or within the spine, application of various types of thermal or radiofrequency energy within the spine, and spinal cord stimulation.

For some interventional therapies, utilization rates have increased dramatically. Rates of epidural steroid injections rose 271% in the Medicare population between 1994 and 2001, and facet joint injections increased 231%.[1] Increased utilization of interventional therapies and surgeries has not been associated with improved health status among patients with low back pain, and may be 1 factor contributing to increases in health care expenditures associated with back pain.[2]

The lack of association between increased utilization of interventional therapies and improved health status represents a challenge to the evidence used to support such treatments. Some interventional procedures have been adopted primarily based on promising case series or other observational studies.[3,4] Such studies can be difficult to interpret due to placebo effects and susceptibility to confounding and bias.[5–8] In addition, a number of interventional therapies target specific areas of the spine thought to be the source of low back pain, but the accuracy of methods for identifying patients with discogenic, facet joint, or sacroiliac joint pain remains uncertain.[9–11]

This article reviews current evidence on benefits and harms of nonsurgical interventional therapies for treatment of low back pain and radiculopathy, focusing on data from randomized controlled trials. It is part of a larger evidence review commissioned by the American Pain Society to guide recommendations for evaluation and management of low back pain.[12]

From the Departments of *Medicine, and †Medical Informatics and Clinical Epidemiology, Oregon Evidence-Based Practice Center, Oregon Health and Science University, Portland, OR; ‡Division of General Medicine, Medical Services, MA General Hospital, Harvard Medical School, Boston, MA; §Department of Physical Medicine and Rehabilitation, Northwestern University, Chicago, IL; and ¶Department of Anesthesiology, University of Iowa, Iowa City, IA.
Acknowledgment date: October 21, 2008. Revision date: December 19, 2008. Acceptance date: December 22, 2008.
The device(s)/drug(s) is/are FDA-approved or approved by corresponding national agency for this indication.
Professional Organization funds were received in support of this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript. Supported by the American Pain Society (APS).
This article is based on research conducted at the Oregon Evidence-Based Practice Center. The authors are solely responsible for the content of this article and the decision to submit for publication. No statement in this article should be construed as an official position of the APS.
Registered Products and Procedures: Coblation, ArthroCare Corporation, 7500 Rialto Boulevard, Building Two, Suite 100, Austin, TX 78735; and Ovid, 333 Seventh Avenue, 20th Floor, New York, NY 1000.
Address correspondence and reprint requests to Roger Chou, MD, 3181 SW Sam Jackson Park Road, Mail code BICC, Portland, OR 97239; E-mail: chour@ohsu.edu

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

Table 1. Included Interventional Therapies

Injections outside the spine	
Botulinum toxin injection	Injection of botulinum toxin (an antispasmodic) into the muscles of the back.
Local injections	Injection of local anesthetic (with or without corticosteroid) into the muscles or soft tissues of the back. Trigger point injections, a type of local injection, involve an injection performed at a tender area, often with a palpable nodule or band.
Prolotherapy	A procedure involving the repeated injection of an irritant chemical into the soft tissues of the back in order to provoke an inflammatory response that will theoretically subsequently lead to strengthening of the soft tissues with decrease in pain and disability. Also referred to as sclerotherapy.
Intraspinal steroid injections and chemonucleolysis	
Chemonucleolysis	Treatment of herniated discs with intradiscal injections of a proteolytic enzyme, most commonly chymopapain (an extract from papaya). Chymopapain acts by digesting the jelly-like inner portion of the disc known as the nucleus pulposus, while at the same time, leaving the outer portion, the annulus fibrosis, essentially intact.
Epidural steroid injection	Injection of corticosteroids via a catheter into the space between the dura and the spine. Common approaches for administering epidural steroid injections are through the interlaminar space, via the neuroforamen under fluoroscopic guidance (transforaminal), and through the sacral hiatus at the sacral canal (caudal).
Facet joint steroid injection	Injection of corticosteroid into the facet joints.
Intradiscal steroid injection	Injection of corticosteroid into the intervertebral disc.
Sacroiliac joint steroid injection	Injection of corticosteroid into or around the sacroiliac joint.
Therapeutic medial branch block	Injection of local anesthetic with or without corticosteroid in the area of the medial branch of the posterior primary ramus, the primary nerve innervating the intervertebral facet joint. Usually used as a diagnostic procedure to identify facet joint pain, but has also been used as a therapeutic procedure.
Radiofrequency denervation, intradiscal electrothermal therapy, and related procedures	
Coblation nucleoplasty	A procedure involving the use of a bipolar radiofrequency current in order to create a series of channels in an intervertebral disc and reduce the volume of tissue.
Intradiscal electrothermal therapy	A procedure involving the placement of an electrode or catheter into the intervertebral disc annulus or nucleus and applying electrothermal energy to alter adjacent pain receptors or other structures.
Percutaneous intradiscal radiofrequency therapy	A procedure involving the placement of an electrode of catheter into the intervertebral disc and applying alternating radiofrequency current. Sometimes classified as a variant of intradiscal electrothermal therapy.
Radiofrequency denervation	A procedure involving the destruction of nerves using heat generated by a radiofrequency current.
Spinal Cord Stimulation	
Spinal cord stimulation	A procedure involving the placement of electrodes in the epidural space adjacent to the area of the spine presumed to be the source of pain and applying an electric current in order to achieve sympatholytic and other neuromodulatory effects.

■ Materials and Methods

Data Sources and Searches

An expert panel convened by the American Pain Society determined which interventional therapies would be included in this review. Table 1 shows the 14 therapies chosen by the panel and how we defined and grouped them.

We conducted searches (through July 2008) combining terms for low back pain with various interventional therapies in Ovid MEDLINE, the Cochrane Database of Systematic Reviews, and the Cochrane Central Register of Controlled Trials (see Appendix 1, Supplemental Digital Content 1, which shows Cochrane and MEDLINE search strategies, http://links.lww.com/A915). Electronic searches were supplemented by reference lists and additional citations suggested by experts. We did not include trials published only as conference abstracts.

Evidence Selection

We included randomized controlled trials and systematic reviews meeting all of the following criteria:

- English language, or non-English language trial but included in an English language systematic review
- Evaluated nonpregnant adults (>18 years old) with low (lumbar or sacral) back pain of any duration, alone or with leg pain
- Evaluated a target injection or other interventional therapy (Table 1)
- Reported at least 1 of the following outcomes: back specific function, generic health status, pain, work disability, or patient satisfaction.[13,14]

We excluded trials of low back pain associated with acute major trauma, cancer, infection, *cauda equina* syndrome, fibromyalgia, spondyloarthropathy, and osteoporosis or vertebral compression fracture. We excluded outdated systematic reviews, which we defined as systematic reviews with a published update, or systematic reviews published before the year 2000. Intrathecal therapy, adhesiolysis, and intradiscal ozone injection are reviewed in the larger report.[12] A separate article addresses surgical interventions for low back pain.[15]

Data Extraction and Quality Assessment

For all placebo-controlled trials and for active-controlled trials not included in previously published systematic reviews, we abstracted information on study design, participant characteristics, interventions, and results. Two reviewers independently rated the quality of these trials using the 11 criteria developed by the Cochrane Back Review Group (see Appendix 2, Supplemental Digital Content 2, http://links.lww.com/A911).[16] Discrepancies were resolved through joint review and a consensus process. We considered trials receiving more than half of the max-

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

imum possible quality score (6 or greater) "higher quality."[16] We did not independently abstract active-controlled trials included in previously published systematic reviews. Instead, we relied on data abstraction and quality ratings as reported by higher quality systematic reviews.

For each included systematic review, we abstracted information on search methods, inclusion criteria, methods for rating study quality, characteristics of included studies, methods for synthesizing data, and results, including the number and quality of trials for each comparison and outcome. We assessed internal validity (quality) of systematic reviews using the Oxman and Guyatt criteria (see Appendix 3, Supplemental Digital Content, http://links.lww.com/A910).[17,18] Using this system, systematic reviews receiving a score of 4 or less (on a scale of 1 to 7) have potential major flaws and are more likely to produce biased conclusions about effectiveness of interventions.[17,19] We classified such systematic reviews "lower quality" and those receiving scores of 5 or more higher quality.

Data Synthesis

We assessed overall strength of evidence for a body of evidence using methods adapted from the US Preventive Services Task Force.[20] To assign an overall strength of evidence (good, fair, or poor), we considered the number, quality, and size of studies; consistency of results between studies; and directness of evidence (see Appendix 4, Supplemental Digital Content 4, for minimum criteria for fair- and good-quality ratings, http://links.lww.com/A914). Consistent results from a number of higher quality studies across a broad range of populations support a high degree of certainty that the results of the studies are true (the entire body of evidence would be considered "good-quality"). For a "fair-quality" body of evidence, results could be due to true effects or to biases operating across some or all of the studies. For a "poor-quality" body of evidence, any conclusion is uncertain.

To evaluate consistency, we classified conclusions of trials and systematic reviews as positive (the intervention is beneficial), negative (the intervention is harmful or not beneficial), or uncertain (imprecise estimates, unclear evidence, or inconsistent results).[17] We defined "inconsistency" as >25% of higher quality trials reaching discordant conclusions (positive *vs.* negative), 2 or more higher quality systematic reviews reaching discordant conclusions, or unexplained heterogeneity (for pooled data). When inconsistency was present, we evaluated differences in how systematic reviews selected or synthesized studies and differences in trial quality, interventions, duration of low back pain, and presence or absence of radiculopathy as potential sources.

We considered mean improvements of 5 to 10 points on a 100-point visual analog pain scale (or equivalent) as small, 10 to 20 points as moderate, and >20 points as large.[21] For back-specific functional status, we classified mean improvements of 2 to 5 points on the Roland Morris Disability Questionnaire (RDQ: 0–24 scale) and 10 to 20 points on the Oswestry Disability Index (ODI: 0–100 scale) as moderate.[22] We also considered standardized mean differences of 0.2 to 0.5 as small, 0.5 to 0.8 as moderate, and >0.8 as large.[23] When reported, we considered a relative risk (RR) of 1.25 to 2.00 for the proportion of patients reporting >30% pain relief (or a similar outcome) a moderate benefit.

Results

Results of Literature Search

The literature search yielded a total of 1331 citations. We retrieved 174 articles based on examination of titles and abstracts. Of 116 full-text articles potentially reporting a relevant randomized controlled trial, we judged 105 to meet inclusion criteria. Of 58 full-text articles potentially reporting a relevant systematic review, we judged 30 (reporting 26 systematic reviews) to meet inclusion criteria (see Appendices 5 and 6, Supplemental Digital Content 5 and 6, http://links.lww.com/A913 and http://links.lww.com/A916).[3,4,24–51] Seventy-five trials (reported in 83 articles) were included in 1 or more previously published systematic reviews (see Appendix 7, Supplemental Digital Content 7, http://links.lww.com/A917). Twenty-two trials not included in any previous systematic review also met inclusion criteria.[52–73] Of 97 total trials, 56 (reported in 56 articles) were placebo-controlled (Appendices 8 and 9, Supplemental Digital Content 8 and 9, http://links.lww.com/A918 and http://links.lww.com/A919).[55–57,59,63–65,68–72,74–120] We excluded 28 potentially relevant reviews[121–148] (Appendix 10, Supplemental Digital Content 10, http://links.lww.com/A920) and 11 trials[149–159] (see Appendix 11, Supplemental Digital Content 11, http://links.lww.com/A921).

Injections Outside the Spine

Local Injections

Four randomized trials evaluated local injections.[82,89,93,116] Two systematic reviews included three[46] or all four[44] of the trials.

Three small (n = 15–41), placebo-controlled trials (1 rated higher quality[82]) each found local anesthetic injections superior to placebo for short-term (≤2 weeks) pain relief for subacute or chronic back pain.[82,93,116] However, the trials evaluated heterogeneous injection methods and patient populations. One trial evaluated a local anesthetic injection over the iliac crest for iliac crest pain,[82] 1 evaluated local anesthetic plus corticosteroid injections over the iliolumbar ligament for nonspecified low back pain,[116] and 1 evaluated local anesthetic trigger point injections for lumbar or cervical (2 of 15 patients) myofascial pain syndrome.[93] None evaluated longer term outcomes. A fourth, higher-quality trial (N = 63) found no difference between a local anesthetic trigger point injection (with or without a corticosteroid) and either a dry needlestick or acupressure for low back strain, but the comparator interventions could be considered active treatments.[89] A higher quality systematic review also found no strong evidence to support local injections.[46]

One lower quality trial found the addition of a corticosteroid to a local anesthetic trigger point injection to be no more effective than a local anesthetic alone.[89] Adverse events reported in local injection trials include pain at the injection site, temporary paresthesia, and nausea.[82,89]

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

Botulinum Toxin Injection

One small (n = 31), higher quality randomized trial of chronic low back pain that found botulinum toxin A injection superior to saline injection for achieving >50% short-term pain relief (73% *vs.* 25% at 3 weeks, P = 0.012; and 60% *vs.* 12% at 8 weeks, P = 0.09) and improvement in ODI scores (67% *vs.* 19%, P = 0.011).[56] Sixty percent of patients with an initial response reported cessation of benefits after 3 to 4 months. Although no adverse events were reported in this trial, case reports exist of allergic reactions, including fatal anaphylaxis.[160]

Prolotherapy

Five randomized, placebo-controlled trials evaluated prolotherapy.[85,97,102,107,119] All were included in a higher quality Cochrane review.[29] We rated 4 trials higher quality.[85,97,107,119]

For chronic nonspecific low back pain, 3 trials (2 higher quality[85,119]) found no difference between prolotherapy and either saline or local anesthetic control injections for short- or long-term (up to 24 months) pain or disability.[102] One higher quality trial found prolotherapy associated with increased likelihood of short-term improvement in pain or disability *versus* control injection (RR = 1.47, 95% CI: 1.04–2.06), but both treatment groups received a number of cointerventions including spinal manipulation, local injections, exercises, and walking.[97] In the fifth trial, effects of prolotherapy could not be determined because the prolotherapy group received strong manipulation and the control injection group only light manipulation.[107] A higher quality Cochrane review rated all 5 placebo-controlled trials of higher quality, and also found prolotherapy to be ineffective when used alone for chronic low back pain.[29]

Nearly all patients in most trials experience the expected temporary increase in back pain and stiffness following prolotherapy injections.[97,107,119] Postinjection headaches suggestive of lumbar puncture occurred in 2% to 4% of patients in 2 trials.[97,119]

■ Intraspinal Steroid Injections and Chemonucleolysis

Epidural Steroid Injection

Forty randomized trials (reported in 40 articles) evaluated epidural steroid injections.[52,53,57,58,62,66,74,76,77,79,81,83,86,94,96,98,99,102,103,109,110,112,115,118,120,161–175] Thirty-three trials were included in at least 1 of 9 systematic reviews[24,27,30,37,39,44,46,49,51] and we identified 7 additional trials.[52,53,57,58,62,66,163] Twenty-one trials (with 2 trials reported in 1 article[99]) were placebo-controlled.[57,74,76,77,79,81,83,86,94,96,98,99,102,103,109,110,112,115,118,120] We rated 9 placebo-controlled trials higher quality.[57,74,79,81,86,96,103,110,118]

For low back pain with radiculopathy, we found inconsistent results for short-term (up to 1 month following injection) benefits, with 10 of 17 trials (including 3 of 7 higher quality trials) showing no differences in pain or function between epidural steroid and placebo injection (Table 2).[74,76,77,79,81,83,86,94,96,98,102,103,109,112,115,118,120] Results were more consistent after trials were stratified according to whether the control intervention was an epidural or nonepidural (soft tissue) injection. Five[74,86,94,109,118] of six[102] trials found an epidural steroid injection associated with short-term benefits compared with a nonepidural (primarily interspinous ligament) placebo injection, including all 3 higher quality trials.[74,86,118] Only two[77,79] of eleven[76,81,83,96,98,103,112,115,120] trials found an epidural steroid injection associated with short-term benefits compared with epidural placebo (saline or local anesthetic). One of the positive trials was rated higher quality,[79] both were small (n = 23 and 35) trials of caudal epidural injections. Three other trials reported mixed or unclear results.[96,98,112] Stratification of trials according to duration of symptoms, use of imaging to confirm presence of prolapsed disc, or study quality did not appear to reduce inconsistency in short-term findings.

Four[77,94,99,110] of 18 trials reported long-term (>3 months) benefits following epidural steroid injection, but 3 of these[77,94,99] were rated lower quality and did not report statistical significance of results. Two[110,176] of seven[74,81,112,115,118] trials found epidural steroid injection associated with lower rates of subsequent surgery compared with placebo injection. Among 4 higher quality trials (total n = 533),[74,81,110,118] only 1 small (n = 55) trial[110] reported this effect.

Three higher quality systematic reviews reached discordant conclusions regarding short-term benefits following epidural steroid injection for sciatica or radiculopathy.[37,39,51] For nonacute (>4 weeks) sciatica, a Cochrane review found no difference between epidural steroid *versus* placebo injection for short-term (<6 weeks) pain relief, but only pooled data from 4 trials (RR = 0.93, 95% CI: 0.79–1.09).[39] A second, qualitative systematic review found no differences between epidural steroid and placebo injections in 5 of 7 trials, including 3 of 4 higher quality trials.[37] The third higher quality systematic review found epidural steroid superior to placebo injection for "improvement in symptoms" (OR = 2.2, 95% CI: 1.0–4.7) for acute or chronic sciatica.[51] Its conclusions may be sensitive to inclusion of a trial reporting an unusually high odds ratio for short-term reduction in inpatient analgesic consumption immediately following the injection (OR = 6.8, compared with 1.1–2.8 in the other trials).[86] Although this trial was excluded from the Cochrane review because it allowed enrollment of patients with acute symptoms, only 10% had symptoms less than 4 weeks. Three lower quality systematic reviews each found some evidence for short-term pain relief following epidural steroid injections for radiculopathy, but also at least some inconsistency between trials.[24,27,49] None of the systematic reviews evaluated results stratified according to use of epidural or soft-tissue placebo injection.

There is insufficient evidence to determine the optimal route of epidural steroid administration. Most placebo-

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

Table 2. Main Results, Placebo-Controlled Trials of Epidural Steroid Injection for Radiculopathy

Study, Yr	Duration	Imaging-Confirmed Prolapsed Disc	Sample Size	Epidural Steroid Injection Administration Route	Control Intervention	Short-Term (<3 mo) Results	Long-Term Results (≥3 mo)	Higher-Quality (Number of Criteria Met/Total Number of Criteria)
Arden *et al*[74]	Subacute or chronic	No	228	Interlaminar	Interspinous ligament saline injection	Positive at 3 wk	Negative at 1 yr	Yes (9/11)
Beliveau[76]	Duration not reported	No	48	Interlaminar	Epidural local anesthetic injection	Negative at 1–3 mo	Not assessed	No (1/11)
Breivik *et al*[77]	Chronic	No	35	Caudal	Epidural local anesthetic plus large volume (100 cc) saline	Positive, but duration of follow-up unclear	Positive, but duration of follow-up unclear	No (5/11)
Bush and Hillier *et al*[79]	Subacute or chronic	No	23	Caudal	Epidural saline injection	Positive at 4 wk (*P* value not reported)	Mixed results at 1 yr	No (6/11)
Carette *et al*[81]	Subacute or chronic	Yes	158	Interlaminar	Epidural saline injection (small volume)	Negative at 3 wk	Negative at 3 mo	Yes (10/11)
Cuckler *et al*[83]	Primarily chronic	No	73	Interlaminar	Epidural local anesthetic injection	Negative at 24 h	Negative at 13–30 mo	No (5/11)
Dilke *et al*[86]	Primarily subacute and chronic	No	100	Interlaminar	Interspinous ligament saline injection	Positive during hospital admission	Negative at 3 mo	Yes (7/11)
Helliwell[94]	Primarily chronic	No	39	Interlaminar	Interspinous ligament saline injection	Positive at 1 mo	Positive at 3 mo (*P* value not reported)	No (2/11)
Karpinnen *et al*[96]	Subacute or chronic	No	160	Transforaminal	Epidural saline injection	Mixed at 2–4 wk	Negative at 1 yr	Yes (10/11)
Klenerman *et al*[98]	Duration not reported (<6 mo by inclusion criteria)	No	73	Interlaminar	Epidural local anesthetic or saline injection	Unclear at 2 mo (*P* value not reported)	Not assessed	No (2/11)
Kraemer *et al*[99]	Duration not reported	Yes	133	Oblique interlaminar	Epidural saline plus intramuscular steroid	Not assessed	Positive at 3 mo (*P* value not reported)	No (2/11)
Kraemer *et al*[99]	Duration not reported	Yes	49	Oblique interlaminar	Epidural saline plus intramuscular steroid	Not assessed	Unclear (*P* value not reported)	No (5/11)
Mathews *et al*[102]	Acute and subacute	No	57	Caudal	Sacral hiatus or tender point local anesthetic injection	Negative at 1 mo	Negative at 1 yr	No (4/11)
Ng *et al*[103]	Chronic	No	88	Transforaminal	Epidural local anesthetic injection	Negative at 6 wk	Negative at 3 mo	Yes (11/11)
Ridley *et al*[109]	Mixed duration	No	39	Interlaminar	Interspinous ligament saline injection	Positive at 2 wk	Not assessed	No (5/11)
Riew *et al*[110,111]	Subacute or chronic	Yes	55	Transforaminal	Epidural local anesthetic injection	Not assessed	Positive at 1 yr (for proportion undergoing surgery)	Yes (9/11)
Rogers *et al*[112]	Chronic	No	30	Interlaminar	Epidural saline injection	Unclear at 1 mo (*P* value not reported)	Negative at 20–21 mo (for proportion undergoing surgery)	No (5/11)
Snoek *et al*[115]	Mixed duration	Yes	51	Interlaminar	Epidural saline injection	Negative at 2 mo	Negative at 8–20 mo	No (4/11)
Wilson-MacDonald *et al*[118]	Subacute or chronic	Yes	92	Interlaminar	Intramuscular and interspinous ligament steroid injection	Positive at 1 mo	Negative at ≥2 yr	Yes (9/11)
Zahaar[120]	Primarily chronic	Yes	63	Caudal	Epidural saline injection	Negative at 24 h	Negative at 20–21 mo (for proportion undergoing surgery)	No (3/11)

controlled trials evaluated the interlaminar or caudal approach. Three higher quality, placebo-controlled trials evaluating the transforaminal approach reported mixed results (Table 2).[96,103,110] No approach was clearly superior in 5 trials (1 rated higher quality[52]) that directly compared different methods.[99,167,168,175] No trial compared fluoroscopic guidance *versus* epidural injection without fluoroscopy. One higher quality trial found that if a first epidural injection was not effective, additional injections within the first 6 weeks were no more effective.[74]

Evidence on efficacy of epidural injections specifically for spinal stenosis,[57,83,120] low back pain without radiculopathy,[174] or failed back surgery syndrome[161,165,169,173] is sparse and inconclusive, but

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

showed no clear benefit.[44,46] One higher quality trial found epidural corticosteroid injection moderately to substantially inferior to discectomy for short-term (1–3 months) outcomes in patients with a large herniated disc, but most differences were no longer observed after 2 to 3 years, and results are difficult to interpret due to high crossover rates and lack of intention-to-treat analysis.[163]

Although there are case reports of serious adverse events, including paralysis and infection, after epidural steroid injection,[177–179] serious adverse events were rarely reported in randomized trials. However, reporting of harms was suboptimal. Ten placebo-controlled trials did not report harm at all.[57,77,83,96,102,109,110,112,118,120] When reported, adverse events were typically transient and minor. In 1 recent high quality trial, 3.3% of 120 patients receiving an interlaminar epidural steroid injection experienced a postinjection headache, 0.8% postdural puncture headache, 1.7% nausea, and 4.2% other adverse events.[74] Serious adverse events were also uncommon in trials that specifically evaluated the transforaminal approach.[96,99,103,167]

Facet Joint Injection and Therapeutic Medial Branch Block

Eight randomized trials evaluated facet joint injection or medial branch block.[67,80,101,180–185] Seven were included in at least 1 of 4 systematic reviews[28,44–46] and we identified 1 additional trial.[67] Two trials (both evaluating facet joint injection) were placebo-controlled.[80,101]

For chronic low back pain with positive response to an uncontrolled facet joint block, a higher quality trial by Carette *et al*[80] (n = 101) found no difference between steroid *versus* saline facet joint injection for experiencing pain relief after either 1 or 3 months (RR = 0.89, 95% CI: 0.65–1.21 and RR = 0.90, 95% CI: 0.69–1.17, respectively). Although a higher proportion of patients in the steroid injection group experienced marked or very marked improvement after 6 months (46% *vs.* 15%, P = .002), half of the steroid injection patients with positive results at 6 months experienced no benefits at earlier time periods. The biologic rationale for such a delayed response is unclear. There was no significant difference in the likelihood of sustained (improvement at 1, 3, and 6 months) improvement (22% *vs.* 10%, P = 0.19). Furthermore, 6-month differences were attenuated after controlling for increased use of cointerventions in the steroid injection group. A second, lower quality trial by Lilius *et al* (n = 109) found no difference in mean pain scores between facet joint intracapsular or pericapsular steroid injection *versus* saline injection.[101] In this trial, criteria for identifying presumed facet joint pain were clinical, and did not include positive response to diagnostic facet joint block.

A higher quality Cochrane review[46] and two[44,45] of 3 lower quality systematic reviews also found no clear benefits associated with facet joint steroid *versus* placebo injection. A third lower quality systematic review found moderate evidence that facet joint injections are associated with short-term improvement.[28] It excluded the trial by Lilius *et al* because it did not use diagnostic facet joint blocks to select patients,[101] classified the trial by Carette *et al* as favoring facet joint injection,[80] classified an active-controlled trial as demonstrating efficacy of facet joint injection because both intervention groups improved compared with baseline,[180] and included evidence from several small (N <100), nonrandomized studies.

No trial evaluated efficacy of therapeutic medial branch block *versus* sham or placebo injection. Two trials (1 higher quality[184]) found no differences between facet joint corticosteroid injection and medial branch block.[67,184] Another lower quality trial found bilateral lumbar facet joint corticosteroid injections plus a home stretching program no more effective than stretching alone for pain or function in patients with presumed lumbar segmental rigidity.[185] No adverse events other than transient local pain at the injection site were reported in the only higher quality, placebo-controlled trial.[80]

Sacroiliac Joint Steroid Injection

One lower quality systematic review on sacroiliac joint injections[36] included 1 small (n = 10) trial,[156] but it did not meet our inclusion criteria because it evaluated patients with spondyloarthropathy. We identified 1 other small (n = 24), higher quality randomized trial of patients without spondyloarthropathy, sacroiliac area pain, and at least 1 physical examination finding for sacroiliac joint pain.[65] It found a periarticular sacroiliac steroid injection substantially superior to local anesthetic injection for improvement in 1-month pain scores. No adverse events were reported.

Intradiscal Steroid Injection

Six randomized trials evaluated intradiscal steroid injection.[55,60,63,70,186–188] Two trials were included in a higher quality Cochrane review[34,35] and we identified 4 additional trials.[55,60,63,70,188] Three trials were placebo-controlled.[55,63,70] All 3 placebo-controlled trials evaluated intradiscal steroid injection for degenerative disc disease.

For chronic low back pain with magnetic resonance imaging evidence of degenerative disc disease and positive response to provocative discography, 2 trials (1 higher quality[70]), found no significant difference between intradiscal steroid and control injections (saline or local anesthetic) for pain relief or improvement in functional status.[63,70] A third, lower quality trial of patients with degenerative disc disease who did not respond to an epidural steroid injection found discography plus intradiscal steroid superior to discography alone only in the subgroup of patients with inflammatory endplate changes on magnetic resonance imaging.[55] However, changes in outcome scores and levels of statistical significance were not well reported.

For back pain with sciatica, 2 trials[186,187] (1 higher quality[186]) included in the Cochrane review found no

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

differences between intradiscal steroid injection and chemonucleolysis for risk of failure or no improvement (OR = 1.20, 95% CI: 0.61–2.38). A third, lower quality trial reported a trend favoring chemonucleolysis over intradiscal steroid injection after 2 years.[60,188] No trial reported on adverse events (such as discitis) following intradiscal steroid injection.

Chemonucleolysis

Twenty-two randomized trials evaluated chemonucleolysis.[61,73,78,84,87,95,113,186,187,189–202] Nineteen trials were included in a higher quality Cochrane review[34,35] and we identified 3 additional trials.[61,73,190] Six trials were placebo-controlled.[78,84,87,95,113,193] We rated 4 of the 5 English-language trials higher quality.[78,84,95,113] A sixth, small (N = 39) French-language trial was included in the Cochrane review.[193]

For lumbar disc prolapse, three[84,87,95] of four[113] English-language trials found chymopapain chemonucleolysis superior to placebo for achieving treatment success (variably defined). A fifth trial found collagenase chemonucleolysis superior to placebo.[78] Based on pooled results, a higher quality Cochrane review that also included a French-language trial[193] found chymopapain chemonucleolysis associated with a lower-likelihood for a poor patient-reported overall outcome ("no success") compared with placebo after 1 year (OR = 0.24, 95% CI: 0.12–0.49, 2 trials), and lower likelihood of open discectomy within 6 to 24 months (OR = 0.41, 95% CI: 0.25–0.68, 5 trials).[34,35] However, compared with initial discectomy, chymopapain chemonucleolysis was associated with a higher likelihood of "unchanged" or "worse" outcomes (OR = 1.64, 95% CI: 0.81–3.33, 2 trials), and a much greater likelihood for subsequent or repeat surgery (OR = 14.29, 95% CI: 5.56–50, 4 trials). About 30% of patients randomized to chemonucleolysis underwent disc surgery within 2 years.

Allergic reactions (including anaphylaxis) occurred in 2%[78,203,204] to 12%[73] of patients undergoing chymopapain chemonucleolysis. Although estimates of allergic reactions vary depending on how they are assessed and defined, severe anaphylaxis appears uncommon. Other serious but rare complications after chemonucleolysis include lumbar subarachnoid hemorrhage and paraplegia.[205,206]

■ Radiofrequency Denervation, Intradiscal Electothermal Therapy (IDET), and Related Procedures

Radiofrequency Denervation

Nine randomized trials evaluated radiofrequency denervation.[59,68,69,71,72,92,100,117,207] Four trials were included in at least 1 of 5 systematic reviews.[28,31,42–45] and we identified 5 additional trials.[59,68,69,71,72] Eight of 9 trials were placebo-controlled.[59,68,69,71,72,92,100,117]

For presumed facet joint pain, trials of radiofrequency denervation are difficult to interpret. The only trial (n = 40) to use controlled facet joint blocks to select patients and an ablation technique believed to be optimal[131] found radiofrequency denervation superior to sham treatment by −1.4 to −1.6 points (0–10 visual analog scale [VAS] scale) for improvement in generalized, back, and leg pain after 6 months, but the difference was not statistically significant for back pain (the main symptom thought to be associated with facet pain).[68] In addition, baseline pain scores in the radiofrequency denervation group averaged 1.6 points higher ($P < 0.05$ for differences) than in the sham group, which suggests inadequate randomization and could be associated with regression to the mean or differential potential for improvement. Furthermore, final pain scores in both groups were identical. Three other trials met criteria to be classified as higher quality but used uncontrolled diagnostic facet joint blocks to select patients, may have used suboptimal techniques,[131,208,209] and reported conflicting results.[72,100,117] One trial (n = 30) found radiofrequency denervation associated with moderately greater improvement in mean VAS pain (−2.4 *vs.* −0.4 on a 0–10 scale, $P < 0.05$) and ODI scores (−11.1 *vs.* +1.7, $P < 0.05$) *versus* sham through 2 months.[117] Radiofrequency denervation was also associated with greater likelihood of experiencing at least a 2 point reduction in VAS pain score and greater than 50% improvement in global effect at 8 weeks (67% *vs.* 37.5%, $P = 0.003$) and 12 months (46.7% *vs.* 12.5%, $P = 0.02$). The second trial (n = 70) found radiofrequency denervation superior to sham treatment for mean improvement in RDQ scores at 4 weeks (−8.4 *vs.* −2.2, $P = 0.05$), but there were no statistically significant differences in ODI or VAS pain scores.[100] At 12 weeks, the difference in RDQ scores was no longer present. The third trial[72] (n = 82) found no differences between radiofrequency and sham intervention on any outcome.

A lower quality trial (n = 60) found conventional but not pulsed radiofrequency denervation superior to sham denervation for pain, the ODI, and analgesic use through 1 year.[71] Effects on pain were small to moderate (0.8–1.5 points on a 0–10 scale) and on the ODI were small (4–6 points). Another sham-controlled trial had serious methodologic shortcomings, including lack of intention-to-treat analysis.[92]

Two higher-quality[31,42,43] and 2 lower quality[44,45] systematic reviews also found uncertain or inconsistent benefits associated with radiofrequency denervation for presumed facet joint pain, though none included the three[68,71,72] most recently published sham-controlled trials. A fifth systematic review concluded there is moderate evidence supporting benefits from radiofrequency denervation.[28] It excluded a higher quality trial[100] with more neutral findings because it used a single block to identify facet joint pain, leaving only a single, small (n = 31) higher quality randomized trial—which also did not appear to use controlled blocks to select patients—demonstrating benefits.[117] This systematic review also included 10 observational studies, but criteria for classifying results of observational studies as positive were poorly described.

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

For presumed discogenic back pain with positive discography unresponsive to treatment with IDET, 1 small (n = 49), lower quality trial found radiofrequency denervation of the ramus communicans nerves associated with substantially better mean VAS pain scores (3.8 *vs.* 6.3 on a 0–10 scale, $P < 0.05$), and moderately better SF-36 bodily pain (43.7 *vs.* 32.4, $P < 0.05$) and physical function scores (58.9 *vs.* 46.5, $P < 0.05$) compared with lidocaine injection after 4 months.[69]

For chronic radicular pain with a positive selective nerve root block, 1 higher quality trial found no difference between radiofrequency denervation of the dorsal root ganglions and sham treatment for achieving clinical success (16% *vs.* 25%, $P = 0.43$), improvement in SF-36 scores, or use of analgesics.[59] There was a trend toward a higher proportion of patients in the sham intervention group that reported >50% reduction in VAS-pain scores for the leg (21% *vs.* 42%, $P = 0.051$).

One trial reported a case of mild, subjective, and transient lower limb weakness following radiofrequency denervation.[69] Two other trials found no difference in adverse events between radiofrequency denervation and sham, though radiofrequency treatment was associated with trends toward increased postprocedural pain.[59,72]

Intradiscal Electrothermal Therapy

Two higher quality, sham-controlled, randomized trials evaluated IDET.[91,108] Five systematic reviews each included 1[26,40] or both[4,25,32,33] of the trials.

For chronic low back pain with positive response to provocative lumbar discography, 2 small (n = 57 and n = 64), higher quality, sham-controlled trials of IDET reported inconsistent results.[91,108] In 1 trial, IDET was associated with moderately greater improvements in mean VAS pain scores (2.4 *vs.* 1.1 on a 0–10 scale, $P = 0.0045$) and slightly greater improvements in mean ODI scores (11 *vs.* 4, $P = 0.050$) compared with sham IDET, but was no better on the SF-36 bodily pain or physical functioning subscales.[108] The proportion of patients with at least a 2-point improvement in VAS pain scores also favored IDET (56% or 18/32 *vs.* 38% or 9/24). The trial appeared to evaluate a highly selected subset of patients, as only 64 patients from a potential cohort of 4253 were enrolled. The other trial found no differences between IDET and sham on the Low Back Pain Outcome Score, ODI, SF-36, or Zung Depression Index.[91]

Two higher quality[4,32,33] and 1 lower quality[40] systematic review also found inconsistent data on efficacy of IDET. Two other lower quality systematic reviews concluded that IDET is effective, largely based on pooled rates of response to IDET from mostly uncontrolled observational studies.[25,26] In 1 controlled observational study included in these reviews, IDET was associated with substantially better VAS pain scores at 3 months (3.5 *vs.* 8.0 on a 0–10 scale, $P < 0.0005$) and 24 months (3.0 *vs.* 7.5, $P = 0.028$), as well as a higher proportion pain-free at 24 months (20% or 7 of 35 *vs.* 0% or 0 of 17).[149]

Transient and mild adverse events following IDET such as radicular pain, paresthesias, and numbness range in incidence from 0% (0 of 58) to 15% (5 of 33).[44] More serious but uncommon or rare adverse events include cerebrospinal fluid leak,[210] *cauda equina* syndrome, and vertebral osteonecrosis.[41]

Percutaneous Intradiscal Radiofrequency Thermocoagulation (PIRFT) and Coblation Nucleoplasty

Two randomized trials evaluated PIRFT.[75,211] Four systematic reviews each included one[32,33,42,212] or both[4,40] of the trials.[4,32,33,40,42,43] A lower quality systematic review of Coblation nucleoplasty identified no randomized trials and insufficient evidence from small case series to evaluate efficacy.[3]

For chronic, presumed discogenic low back pain based on a positive response to analgesic discography (N = 28), 1 higher quality placebo-controlled trial[75] found no significant differences between PIRFT and sham PIRFT in improvement in VAS pain scores, global effect, ODI, or a composite outcome of overall treatment success.[75] A second trial was not sham-controlled, but found minimal improvement with either lower- or higher-intensity of PIRFT.[211] Discitis was reported as a complication in 1 trial.[211]

■ Spinal Cord Stimulation

Two randomized trials evaluated spinal cord stimulation.[64,104] Interim or final results of 1 trial[104] were included in 3 higher quality systematic reviews[38,47,48,50] and we identified 1 additional trial.[64]

For failed back surgery syndrome with persistent radiculopathy, 1 higher quality trial (N = 50) found spinal cord stimulation associated with a greater likelihood for >50% pain relief compared with reoperation after a mean of 2.9 years (38% or 9 of 24 *vs.* 12% or 3 of 26, $P = 0.048$).[104] Spinal cord stimulation was also associated with a lower rate of increased use of opioids (13% *vs.* 42%), and fewer patients allocated to spinal cord stimulation subsequently underwent surgery (21% or 5 of 24) compared with those allocated to surgery who later received spinal cord stimulation (54% or 14 of 26). Three-year results were similar. A second, higher quality trial (N = 100) of patients with persistent radicular pain following anatomically successful surgery for herniated disc found spinal cord stimulation associated with greater likelihood of experiencing >50% pain relief after 6 months compared with conventional medical management (48% *vs.* 9%, $P < 0.001$).[64] Spinal cord stimulation was also moderately superior (by 10–20 points) on 7 of 8 SF-36 subscales and the ODI. The trial was designed so that patients randomized to spinal cord stimulation would undergo device implantation only if they experienced greater than 50% pain relief or 80% paresthesia coverage following a screening trial. Ninety-two percent (48 of 52) of patients randomized to spinal cord stimulation underwent implantation, including 5 pa-

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

Table 3. Summary of Evidence for Nonspinal Injections, Intraspinal Steroid Injections, and Chemonucleolysis

Intervention	Population	No. Placebo-Controlled Trials (No. Rated Higher-Quality)	Placebo-Controlled Trials With ≥100 Patients	Total No. Trials	Net Benefit *vs.* Placebo*	Inconsistency†	Directness of Evidence	Overall Quality of Evidence	Comments
Nonspinal injections									
Botulinum toxin injection	Nonspecific low back pain	1 (1)	0	1	Moderate (short-term only, 1 small trial)	Not applicable	Direct	Poor	
Local injections	Nonspecific low back pain	3 (1)	0	5	Unable to determine	No	Direct	Poor	Interventions and populations varied substantially between trials. No higher-quality trials, all trials had small sample sizes
Prolotherapy	Nonspecific low back pain	5 (4)	1	5	No effect	No	Direct	Good	
Intraspinal steroid injections and chemonucleolysis									
Chemonucleolysis	Radiculopathy with prolapsed lumbar disc	6 (5)‡	2	22	Moderate	No	Direct	Good	Chemonucleolysis with chymopapain superior to placebo injection, but inferior to surgery
Epidural steroid injection	Radiculopathy with prolapsed lumbar disc	21 (9)	5	34	Moderate (short-term only)	Yes	Direct	Fair	Inconsistency between higher-quality trials could be due to use of epidural or nonepidural placebo injection
	Spinal stenosis	3 (1)	0	3	Unable to determine	No	Direct	Poor	In 2 of 3 trials, only a subgroup of patients had spinal stenosis
	Non-specific low back pain	0	Not applicable	1	No evidence	Not applicable	Direct	Poor	No difference between epidural steroid and intrathecal midazolam injection in 1 small trial
	Failed back surgery syndrome	0	Not applicable	4	No evidence	No	Direct	Poor	
Facet joint steroid injection	Presumed facet joint pain	2 (1)	2	7	No effect	No	Direct	Fair	
Intradiscal steroid injection	Radicul-opathy with prolapsed lumbar disc	0	Not applicable	3	No evidence	No	Direct	Fair	No effect *vs.* chemonucleolysis
	Presumed discogenic low back pain	3 (1)	2	3	No effect	No	Direct	Good	
Medial branch block (therapeutic)	Presumed facet joint pain	0	Not applicable	3	No evidence	No	Direct	Poor	
Sacroiliac joint steroid injection	Presumed sacroiliac joint pain	1 (1)	0	1	Substantial (1 small trial)	Not applicable	Direct	Poor	The only available trial evaluated a periarticular corticosteroid injection

*Based on evidence showing intervention is more effective than placebo or sham therapy for 1 or more of the following outcomes: pain, functional status, overall improvement, or work status *versus* placebo, small benefit defined as 5–10 points on a 100-point VAS for pain (or equivalent), 1–2 points on the RDQ, 5–10 points on the ODI, or a standardized mean difference (SMD) of 0.2–0.5. Moderate benefit defined as10–20 points on a VAS for pain, 2–5 points on the RDQ, 10–20 points on the ODI, or a SMD of 0.5–0.8. Large benefit defined as >20 points on a 100 point VAS for pain, >5 points on the RDQ, >20 points on the ODI, or a SMD of >0.8.
† Inconsistency defined as <75% of trials reaching consistent conclusions on efficacy (no effect *versus* positive effect considered inconsistent).
‡Quality of 1 small French-language trial not assessed.

tients who did not meet criteria for a positive trial. In the randomized trials, 26% to 32% of patients experienced a complication following spinal cord stimulator implantation, including electrode migration, infection or wound breakdown, generator pocket-related complications, and lead problems.[64,104]

No randomized trials evaluated spinal cord stimulation for chronic low back pain not related to the failed back surgery syndrome or for failed back surgery syndrome without radiculopathy. Although a systematic review identified 27 case series of spinal cord stimulation for chronic low back pain not related to the failed back surgery syndrome that did not meet our inclusion criteria, the methodologic quality of these studies was very low (median quality score: 1 of maximum 7).[48]

■ Discussion

This review synthesizes evidence from randomized controlled trials of injections and other nonsurgical interventional therapies for low back pain and radiculopathy. Main results are summarized in Table 3 and 4. Among the 14 therapies included in this review, we found few supported by even fair quality evidence of benefits from randomized,

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

Table 4. Summary of Evidence for Spinal Cord Stimulation, Intradiscal Electrothermal Therapy, Radiofrequency Denervation, and Related Interventions

Intervention	Population	No. Placebo-Controlled Trials (No. Rated Higher-Quality)	Placebo-Controlled Trials With ≥100 Patients	Total No. Trials	Not Benefit *vs.* Placebo*	Inconsistency†	Directness of Evidence	Overall Quality of Evidence	Comments
Coblation nucleoplasty	Radiculopathy with prolapsed lumbar disc	0	Not applicable	0	No evidence	Not applicable	Not applicable	Not applicable	
	Presumed discogenic low back pain	0	Not applicable	0	No evidence	Not applicable	Not applicable	Not applicable	
IDET	Presumed discogenic low back pain	2 (2)	0	2	Unable to determine (2 trials with inconsistent results)	Yes	Direct	Poor	
PIRFT	Presumed discogenic low back pain	1 (1)	0	2	No benefit (1 trial)	Not applicable	Direct	Poor	
Radiofrequency denervation	Radiculopathy with prolapsed lumbar disc	1 (1)	0	1	No benefit (1 trial)	Not applicable	Direct	Poor	
	Presumed facet joint pain	6 (4)	0	6	Unable to determine	Yes	Direct	Poor	1 higher-quality trial used an inadequate technique, another had large baseline differences in pain scores
	Presumed discogenic low back pain	1 (0)	0	1	Unable to determine (1 trial)	Not applicable	Direct	Poor	
Spinal cord stimulation	Failed back surgery syndrome with persistent radiculopathy	1 (1)	0	2	Moderate (see comments)	No	Direct	Fair	Spinal cord stimulation superior to repeat surgery in 1 trial and superior to conventional medical management in a second trial
	Nonspecific low back pain, or radiculopathy with prolapsed lumbar disc	0	Not applicable	No trials	No evidence	Not applicable	Not applicable	Not applicable	

*Based on evidence showing intervention is more effective than placebo or sham therapy for 1 or more of the following outcomes: pain, functional status, overall improvement, or work status *versus* placebo, small benefit defined as 5–10 points on a 100-point VAS for pain (or equivalent), 1–2 points on the RDQ, 5–10 points on the ODI, or a SMD of 0.2–0.5. Moderate benefit defined as10–20 points on a VAS for pain, 2–5 points on the RDQ, 10–20 points on the ODI, or a SMD of 0.5–0.8. Large benefit defined as >20 points on a 100 point VAS for pain; >5 points on the RDQ, >20 points on the ODI, or a SMD of >0.8.
†Inconsistency defined as <75% of trials reaching consistent conclusions on efficacy (no effect *versus* positive effect considered inconsistent).

placebo-controlled trials. Some interventional therapies were supported by evidence of moderate benefits, but conclusions were qualified by factors such as inconsistency between trials, lack of long-term benefits, relatively sparse data, or limited availability in the United States.

We found good evidence that chymopapain chemonucleolysis is moderately effective compared with placebo injection for sciatica or radiculopathy with prolapsed lumbar disc. However, chemonucleolysis is inferior to initial surgical discectomy, and a high proportion of patients who receive chemonucleolysis subsequently undergo surgery.[34,35] Chemonucleolysis is no longer widely practiced or available in the United States, partly due to concerns about allergic reactions, though the incidence of serious allergic reactions appears to be low.

For epidural steroid injection, we found fair evidence of moderate benefit compared with placebo injection for short-term pain relief in patients with radiculopathy. There was no evidence of long-term benefits, but few trials evaluated long-term outcomes. For short-term outcomes, inconsistency was present between trials. The inconsistency may be related to the type of control used. Specifically, trials that evaluated a soft-tissue placebo injection more consistently reported short-term benefits, and trials that evaluated an epidural placebo injection mostly reported no short-term benefits. This observation suggests that effects could be mediated more by nonspe-

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

cific physical effects of increased volume within the epidural space than by specific corticosteroid anti-inflammatory effects.[213,214] However, this hypothesis does not explain results of a higher quality trial that found no difference between epidural steroid injection and a minimal volume (1 mL) epidural placebo injection.[81] Trials that compare epidural saline or local anesthetic injection *versus* a nonepidural placebo injection would be helpful for clarifying the efficacy of nonsteroid epidural injections. Though data are sparse and inconclusive, the few trials that have been published have not shown that epidural steroid injections are effective for spinal stenosis[57,83,120] or for low back pain without radiculopathy.[174]

We found fair evidence from 2 trials that spinal cord stimulation is more effective than either repeat surgery[104] or continued conventional medical management[64] for failed back surgery syndrome with persistent radiculopathy. Unlike the other interventional therapies included in this review, spinal cord stimulation involves the permanent placement of a device and is associated with a high rate of postimplant complications, though these events are usually not serious. No randomized trials evaluated spinal cord stimulation for failed back surgery syndrome without radiculopathy or back pain not related to the failed back surgery syndrome, though a number of case series have been published. However, in addition to use of an uncontrolled study design, the methodologic quality of these studies was very low.[48]

We found good evidence that prolotherapy is ineffective for nonspecific low back pain[29] and intradiscal corticosteroid injection[34,35] is ineffective for presumed discogenic back pain. We also found fair evidence that intra-articular facet joint steroid injection is not effective, though evidence is limited to 2 placebo-controlled trials that did not use controlled diagnostic facet joint blocks to select patients.[80,101] This could have resulted in underestimates of efficacy—assuming that controlled diagnostic facet joint blocks are truly more accurate than uncontrolled blocks for identifying "true" facet joint pain.[9] In addition, interpretation of positive long-term results from the highest quality trial remains controversial, as it is not known why benefits of facet joint injection would only become apparent 6 months following the procedure.[80]

For 3 procedures involving application of different types of energy to degenerated discs for treatment of presumed discogenic back pain, we found no clear evidence of benefits. We were unable to draw reliable conclusions about efficacy of IDET because the 2 available trials provide conflicting results, we found fair evidence that PIRFT thermocoagulation is not effective, and there are no randomized trials of Coblation nucleoplasty.[4] We found insufficient evidence from randomized trials to reach reliable conclusions regarding other interventional therapies, due to conflicting results (radiofrequency denervation for presumed facet joint pain),[42,43] sparse data (botulinum toxin[56] and sacroiliac joint injection[65]), lack of placebo-controlled trials (therapeutic medial branch block), or methodologic shortcomings and clinical heterogeneity (local injections).[39] It is not clear if clinical trials of interventions targeting specific anatomic sources of back pain have failed to demonstrate efficacy because of inaccurate diagnostic methods, because the interventions truly do not work, or (in the case of radiofrequency denervation[131]) because the trials evaluated technically inadequate procedures.

Our evidence synthesis has several potential limitations. We only included randomized controlled trials. Observational studies can provide important insight into effectiveness of interventions in real-world practice, but are also more susceptible to confounding and bias, which can be particularly problematic in studies that evaluate subjective outcomes such as pain. In addition, observational studies of interventional therapies have sometimes shown substantial benefits[149] that have not been confirmed by subsequent randomized controlled trials.[4] Language bias could have affected our results because we only included non-English language trials included in English language systematic reviews. However, we found no additional non-English trials in our searches.

In addition to sparse evidence for several interventional therapies, another factor that limited our ability to reach more definitive conclusions was that most trials assessed or reported harms data poorly, or did not mention harms at all. This is an important shortcoming because all interventional therapies have some potential for complications, some of which may be serious. Better assessment and reporting of harms in clinical trials would help provide more balanced assessments of net benefits.[215]

A strength of our study is that we did not rely on previously published systematic reviews to generate conclusions. Rather, we conducted our own searches and independently abstracted and reviewed placebo-controlled trials. With the exception of epidural steroids, we found higher quality systematic reviews of the same intervention to reach conclusions consistent with our own analyses as well as with each other. As in previous studies, lower quality reviews tended to reached more favorable conclusions regarding efficacy.[17,19] We found that these discrepancies could be explained by methods used by the lower quality systematic reviews, including higher weight placed on observational studies, less weight placed on inconsistency between trials, inconsistent application of inclusion criteria, and more favorable interpretations of unclear trial results.

In summary, evidence from randomized, placebo-controlled trials showing benefits of most nonsurgical interventional therapies for back pain is limited. More evidence is needed to demonstrate efficacy of non-spinal injections (local injections, prolotherapy, and botulinum toxin) and interventional therapies that target presumed facet joint, discogenic, or sacroiliac joint pain. For herniated lumbar disc with radiculop-

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

athy, there is fair evidence of benefits associated with epidural steroid injections and good evidence for chemonucleolysis. The decision to use chemonucleolysis or epidural steroid injection should take into account the short-term nature of symptom relief, inconsistent results of epidural steroid trials, the relative unavailability of chemonucleolysis in the United States, and surgical[216] as well as continued medical management as alternative treatment options. There is also fair evidence of benefits associated with spinal cord stimulation for failed back surgery syndrome with persistent radiculopathy. However, spinal cord stimulation involves the permanent placement of a device and is associated with frequent but usually nonserious postimplantation complications. Given increasing utilization rates, high costs, and potential harms associated with nonsurgical interventional therapies, more well-designed randomized trials are urgently needed to guide their appropriate use.

■ Key Points

- For sciatica or prolapsed lumbar disc with radiculopathy, there is good evidence from randomized trials that chemonucleolysis is moderately superior to placebo injection but inferior to discectomy and fair evidence (some inconsistency among higher quality trials) that epidural steroid injection is moderately effective for short-term (but not long-term) symptom relief.
- There is fair evidence from randomized trials that spinal cord stimulation is moderately effective for failed back surgery syndrome with persistent radiculopathy, though device-related complications are common.
- There is good or fair evidence from randomized trials that prolotherapy, facet joint injection, intradiscal steroid injection, and PIRFT thermocoagulation are not effective.
- There is insufficient (poor) evidence from randomized trials (conflicting trials, sparse and lower quality data, or no randomized trials) to reliably evaluate other interventional therapies, including local injections, botulinum toxin injection, therapeutic medial branch block, sacroiliac joint injection, radiofrequency denervation, IDET, and coblation nucleoplasty.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text, and links to the digital files are provided in the HTML text of this article on the journal's Web site (www.spinejournal.com).

Acknowledgments

The authors thank Laurie Hoyt Huffman and Tracy Dana for data abstraction, and Jayne Schablaske and Michelle Pappas for administrative support with this manuscript.

References

1. Friedly J, Chan L, Deyo R. Increases in lumbosacral injections in the Medicare population. *Spine* 2007;32:1754–60.
2. Martin BI, Deyo RA, Mirza SK, et al. Expenditures and health status among adults with back and neck problems. *JAMA* 2008;299:656–64.
3. NICE (National Institute for Clinical Excellence). *Interventional Procedures Overview of Percutaneous Disc Decompression Using Coblation for Lower Back Pain*. London, UK: National Institute for Clinical Excellence; 2004. IP Guidance No. IPG173.
4. Urrutia G, Kovacs F, Nishishinya MB, et al. Percutaneous thermocoagulation intradiscal techniques for discogenic low back pain. *Spine* 2007;32:1146–54.
5. Bogduk N. A narrative review of intra-articular corticosteroid injections for low back pain. *Pain Med* 2005;6:287–96.
6. Hartz A, Benson K, Glaser J, et al. Assessing observational studies of spinal fusion and chemonucleolysis. *Spine* 2003;28:2268–75.
7. Hoffman RM, Turner JA, Cherkin DC, et al. Therapeutic trials for low back pain. *Spine* 1994;19(18S):2068S–75S.
8. Turner JA, Deyo RA, Loeser JD, et al. The importance of placebo effects in pain treatment and research. *JAMA* 1994;271:1609–14.
9. Bogduk N. Diagnosing lumbar zygapophysial joint pain. *Pain Med* 2005;6:30–3.
10. Carragee EJ, Alamin TF, Carragee JM. Low-pressure positive discography in subjects asymptomatic of significant low back pain illness. *Spine* 2006;31:505–9.
11. van der Wurff P, Meyne W, Hagmeijer RH. Clinical tests of the sacroiliac joint. A systematic methodological review, part 2: validity. *Man Ther* 2000;5:89–96.
12. Chou R, Huffman L. *Evaluation and Management of Low Back Pain: Evidence Review*. Glenview, IL: American Pain Society; 2009.
13. Bombardier C. Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. *Spine* 2000;25:3100–3.
14. Deyo RA, Battie M, Beurskens AJ, et al. Outcome measures for low back pain research: a proposal for standardized use. *Spine* 1998;23:2003–13.
15. Chou R, Baisden J, Carragee E, et al. Surgery for low back pain: a review of the evidence for an American Pain Society clinical practice guideline. *Spine* 2009;34:1094–109.
16. van Tulder M, Furlan AD, Bombardier C, et al; the Editorial Board of the Cochrane Collaboration Back Review Group. Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group. *Spine* 2003;28:1290–9.
17. Furlan AD, Clarke J, Esmail R, et al. A critical review of reviews on the treatment of chronic low back pain. *Spine* 2001;26:E155–62.
18. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol* 1991;44:1271–8.
19. Jadad AR, McQuay HJ. Meta-analyses to evaluate analgesic interventions: a systematic qualitative review of their methodology. *J Clin Epidemiol* 1996;49:235–43.
20. Harris RP, Helfand M, Woolf SH, et al. Current methods of the US Preventive Services Task Force. *Am J Prev Med* 2001;20(3S):21–35.
21. Ostelo RW, Deyo RA, Stratford P, et al. Interpreting change scores for pain and functional status in low back pain. *Spine* 2008;33:90–4.
22. Bombardier C, Hayden JA, Beaton DE. Minimal clinically important difference. Low back pain: outcome measures. *J Rheumatol* 2001;28:431–8.
23. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.
24. Abdi S, Datta S, Trescot AM, et al. Epidural steroids in the management of chronic spinal pain: a systematic review. *Pain Physician* 2007;10:185–212.
25. Andersson GB, Mekhail NA, Block JE. Treatment of intractable discogenic low back pain. A systematic review of spinal fusion and intradiscal electrothermal therapy (IDET). *Pain Physician* 2006;9:237–48.
26. Appleby D, Andersson G, Totta M. Meta-analysis of the efficacy and safety of intradiscal electrothermal therapy (IDET). *Pain Med* 2006;7:308–6.
27. Armon C, Argoff CE, Samuels J, et al. Assessment: use of epidural steroid injections to treat radicular lumbosacral pain: report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. *Neurology* 2007;68:723–9.
28. Boswell MV, Colson JD, Sehgal N, et al. A systematic review of therapeutic facet joint interventions in chronic spinal pain. *Pain Physician* 2007;10:229–53.
29. Dagenais S, Yelland MJ, Del Mar C, et al. Prolotherapy injections for chronic low-back pain. *Cochrane Database Syst Rev* 2007:CD004059.
30. DePalma MJ, Bhargava A, Slipman CW. A critical appraisal of the evidence for selective nerve root injection in the treatment of lumbosacral radiculopathy. *Arch Physical Med Rehabil* 2005;86:1477–83.

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

31. Geurts J, van Wijk R, Stolker R, et al. Efficacy of radiofrequency procedures for the treatment of spinal pain: a systematic review of randomized clinical trials. *Reg Anesth Pain Med* 2001;26:394–400.
32. Gibson J, Waddell G. Surgery for degenerative lumbar spondylosis: updated Cochrane review. *Spine* 2005;30:2312–20.
33. Gibson JNA, Waddell G, Grant IC. Surgery for degenerative lumbar spondylosis. *Cochrane Database Syst Rev* 2005:CD001352.
34. Gibson JN. Surgical interventions for lumbar disc prolapse: updated Cochrane Review. *Spine* 2007;32:1735–47.
35. Gibson J, Waddell G. Surgical interventions for lumbar disc prolapse. *Cochrane Database Syst Rev* 2007:CD001350.
36. Hansen H, McKenzie-Brown AM, Cohen SP, et al. Sacroiliac joint interventions: a systematic review. *Pain Physician* 2007;10:165–84.
37. Luijsterburg PAJ, Verhagen AP, Ostelo RWJG, et al. Effectiveness of conservative treatments for the lumbosacral radicular syndrome: a systematic review. *Eur Spine J* 2007;16:881–99.
38. Mailis-Gagnon A, Furlan A, Sandoval J, et al. Cochrane Pain Palliative Care and Supportive Care Group. Spinal cord stimulation for chronic pain. *Cochrane Database Syst Rev* 2004:CD003783.
39. Nelemans P, deBie R, deVet H, et al. Injection therapy for subacute and chronic benign low back pain. *Spine* 2001;26:501–15.
40. NICE (National Institute for Clinical Excellence). *Interventional Procedures Overview of Percutaneous Intradiscal Radiofrequency Thermocoagulation for Lower Back Pain*. London, UK: National Institute for Clinical Excellence; 2004. IP Guidance Number 83.
41. NICE (National Institute for Clinical Excellence). *Interventional Procedures Overview of Percutaneous Intradiscal Electrothermal Therapy for Lower Back Pain*. London, UK: National Institute for Clinical Excellence; 2004. IP Guidance Number 81.
42. Niemisto L, Kalso E, Malmivaara A, et al. Radiofrequency denervation for neck and back pain. A systematic review of randomized controlled trials. *Cochrane Database Syst Rev* 2003:CD004058.
43. Niemisto L, Kalso E, Malmivaara A, et al. Cochrane Collaboration Back Review Group. Radiofrequency denervation for neck and back pain: a systematic review within the framework of the Cochrane Collaboration Back Review Group. *Spine* 2003;28:1877–88.
44. Resnick DK, Choudhri TF, Dailey AT, et al. Guidelines for the performance of fusion procedures for degenerative disease of the lumbar spine, part 13: injection therapies, low-back pain, and lumbar fusion. *J Neurosurg Spine* 2005;2:707–15.
45. Slipman CW, Bhat AL, Gilchrist RV, et al. A critical review of the evidence for the use of zygapophysial injections and radiofrequency denervation in the treatment of low back pain. *Spine J* 2003;3:310–6.
46. Staal J, de Bie R, De Vet H, et al. Injection therapy for subacute and chronic low-back pain. *Cochrane Database Syst Rev* 2008:CD001824.
47. Taylor RS. Spinal cord stimulation in complex regional pain syndrome and refractory neuropathic back and leg pain/failed back surgery syndrome: results of a systematic review and meta-analysis. *J Pain Symptom Manage* 2006;31(suppl 1):S13–9.
48. Taylor RS, Van Buyten J, Buscher E. Spinal cord stimulation for chronic back pain and leg pain and failed back surgery syndrome: a systematic review and analysis of progressive factors. *Spine* 2005;30:152–60.
49. Tonkovich-Quaranta L, Winkler S. Use of epidural corticosteroids in low back pain. *Ann Pharmacother* 2000;34:1165–72.
50. Turner JA, Loeser JD, Deyo RA, et al. Spinal cord stimulation for patients with failed back surgery syndrome or complex regional pain syndrome: a systematic review of effectiveness and complications. *Pain* 2004;108:137–47.
51. Vroomen P, de Krom M, Slofstra P, et al. Conservative treatment of sciatica: a systematic review. *J Spinal Disord* 2000;13:463–9.
52. Ackerman WE III, Ahmad M. The efficacy of lumbar epidural steroid injections in patients with lumbar disc herniations. *Anesth Analg* 2007;104: 1217–22.
53. Bonetti M, Fontana A, Cotticelli B, et al. Intraforaminal O2–O3 versus periradicular steroidal infiltration in lower back pain: randomized controlled study. *Am J Neuroradiol* 2005;26:996–1000.
54. Burton AK, Tillotson KM, Cleary J. Single-blind randomised controlled trial of chemonucleolysis and manipulation in the treatment of symptomatic lumbar disc herniation. *Eur Spine J* 2000;9:202–07.
55. Buttermann GR. The effect of spinal steroid injections for degenerative disc disease. *Spine J* 2004;4:495–505.
56. Foster L, Clapp L, Erickson M, et al. Botulinum toxin A and chronic low back pain: a randomized, double-blind study. *Neurology* 2001;56:1290–3.
57. Fukusaki M, Kobayashi I, Hara T, et al. Symptoms of spinal stenosis do not improve after epidural steroid injection. *Clin J Pain* 1998;14:148–51.
58. Gallucci M, Limbucci N, Zugaro L, et al. Sciatica: treatment with intradiscal and intraforaminal injections of steroid and oxygen-ozone versus steroid only. *Radiology* 2007;242:907–13.
59. Geurts J, van Wijk RM, Wynne HJ, et al. Radiofrequency lesioning of dorsal root ganglia for chronic lumbosacral radicular pain: a randomised, double-blind, controlled trial. *Lancet* 2003;361:21–6.
60. Graham C. Chemonucleolysis. A preliminary report on a double blind study comparing chemonucleolysis and intradiscal administration of hydrocortisone in the treatment of backache and sciatica. *Orthop Clin North Am* 1975;6:259–63.
61. Hoogland T, Schubert M, Miklitz B, et al. Transforaminal posterolateral endoscopic discectomy with or without the combination of a low-dose chymopapain: a prospective randomized study in 280 consecutive cases. *Spine* 2006;31:E890–7.
62. Jeong HS, Lee JW, Kim SH, et al. Effectiveness of transforaminal epidural steroid injection by using a preganglionic approach: a prospective randomized controlled study. *Radiology* 2007;245:584–90.
63. Khot A, Bowditch M, Powell J, et al. The use of intradiscal steroid therapy for lumbar spinal discogenic pain: a randomized controlled trial. *Spine* 2004;29:833–6.
64. Kumar K, Taylor RS, Jacques L, et al. Spinal cord stimulation versus conventional medical management for neuropathic pain: a multicentre randomised controlled trial in patients with failed back surgery syndrome. *Pain* 2007;132:179–88.
65. Luukkainen R, Wennerstrand P, Kautiainen H, et al. Efficacy of periarticular corticosteroid treatment of the sacroiliac join in nonspondyloarthropathic patients with chronic low back pain in the region of the sacroiliac joint. *Clin Exp Rheumatol* 2002;20:52–4.
66. Manchikanti L, Rivera JJ, Pampati V, et al. One day lumbar epidural adhesiolysis and hypertonic saline neurolysis in treatment of chronic low back pain: a randomized, double-blind trial. *Pain Physician* 2004;7:177–86.
67. Nash TP. Facet joints - intra-articular steroids or nerve block? *Pain Clinic* 1989;3:77–82.
68. Nath S, Nath C, Pettersson K. Percutaneous lumbar zygapophysial (facet) joint neurotomy using radiofrequency current, in the management of chronic low back pain. *Spine* 2008;33:1291–7.
69. Oh WS, Shim JC. A randomized controlled trial of radiofrequency denervation of the ramus communicans nerve for chronic discogenic low back pain. *Clin J Pain* 2004;20:55–60.
70. Simmons JW, McMillin JN, Emery SF, et al. Intradiscal steroids. A prospective double-blind clinical trial. *Spine* 1992;17(suppl 6):S172–5.
71. Tekin I, Mirzai H, OK G, et al. A comparison of conventional and pulsed radiofrequency denervation in the treatment of chronic facet joint pain. *Clin J Pain* 2007;23:524–9.
72. van Wijk R, Geurts J, Wynne H, et al. Radiofrequency denervation of lumbar facet joints in the treatment of chronic low back pain: a randomized, double-blind, sham lesion-controlled trial. *Clin J Pain* 2005;21: 335–44.
73. Wittenberg RH, Oppel S, Rubenthaler FA, et al. Five-year results from chemonucleolysis with chymopapain or collagenase. *Spine* 2001;26: 1835–41.
74. Arden NK, Price C, Reading I, et al. A multicentre randomized controlled trial of epidural corticosteroid injections for sciatica. *Rheumatology* 2005; 44:1399–406.
75. Barendse GA, van Den Berg SG, Kessels AH, et al. Randomized controlled trial of percutaneous intradiscal radiofrequency thermocoagulation for chronic discogenic back pain. *Spine* 2001;26:287–92.
76. Beliveau P. A comparison between epidural anaesthesia with and without corticosteroid in the treatment of sciatica. *Rheumatol Phys Med* 1971;11: 40–3.
77. Breivik H, Hesla PE, Molnar I, et al. Treatment of chronic low back pain and sciatica: comparison of caudal epidural injections of bupivicaine and methylprednisolone with bupivicaine followed by saline. *Adv Pain Res Ther* 1976;1:927–32.
78. Bromley J, Varma A, Santoro A, et al. Double-blind evaluation of collagenase injections for herniated lumbar discs. *Spine* 1984;9:486–8.
79. Bush K, Hillier S. A controlled study of caudal epidural injections of triamcinolone plus procaine for the management of intractable sciatica. *Spine* 1991;16:572–5.
80. Carette S, Marcoux S, Truchon R, et al. A controlled trial of corticosteroid injections into facet joints for chronic low back pain. *N Engl J Med* 1991; 325:1002–7.
81. Carette S, Leclaire R, Marcoux S, et al. Epidural corticosteroid injections for sciatica due to herniated nucleus pulposus. *N Engl J Med* 1997;336:1634–40.
82. Collee G, Dijkmans B, Vandenbroucke J, et al. Iliac crest pain syndrome in low back pain. A double blind, randomized study of local injection therapy. *J Rheumatol* 1991;18:1060–3.

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

83. Cuckler J, Bernini P, Wiesel S, et al. The use of epidural steroids in the treatment of lumbar radicular pain. *J Bone Joint Surg* 1985;67A:63–6.
84. Dabezies E, Langford K, Morris J, et al. Safety and efficacy of chymopapain (Discase) in the treatment of sciatica to a herniated nucleus pulposus. Results of a randomized, double-blind study. *Spine* 1988;13:561–5.
85. Dechow E, Davies R, Carr A, et al. A randomized, double-blind, placebo-controlled trial of sclerosing injections in patients with chronic low back pain. *Rheumatology* 1999;38:1255–9.
86. Dilke TFW, Burry HC, Grahame R. Extradural corticosteroid injection in management of lumbar nerve root compression. *BMJ* 1973;16:635–7.
87. Fraser R. Chymopapain for the treatment of intervertebral disc herniation. A preliminary report of a double-blind study. *Spine* 1982;7:608–12.
88. Fraser R. Chymopapain for the treatment of intervertebral disc herniation. The final report of a double-blind study. *Spine* 1984;9:815–8.
89. Garvey T, Marks M, Wiesel S. A prospective, randomized, double-blind evaluation of trigger-point injection therapy for low-back pain. *Spine* 1989; 14:962–4.
90. Gogan W, Fraser R. Chymopapain: a 10-year, double-blind study. *Spine* 1992;17:388–94.
91. Freeman BJ, Fraser RD, Cain CM, et al. A randomized, double-blind, controlled trial: intradiscal electrothermal therapy versus placebo for the treatment of chronic discogenic low back pain. *Spine* 2005;30:2369–77.
92. Gallagher J, Petriccione D, Wedley J, et al. Radiofrequency facet joint denervation in the treatment of low back pain: a prospective controlled double-blind study to assess its efficacy. *Pain Clinic* 1994;7:193–8.
93. Hameroff S, Crago B, Blitt C, et al. Comparison of bupivacaine, etidocaine, and saline for trigger-point therapy. *Anaesth Analg* 1981;60:752–5.
94. Helliwell M. Outpatient treatment of low back pain and sciatica by a single extradural corticosteroid injection. *Br J Clin Pract* 1985;39:228–31.
95. Javid M, Nordby E, Ford L, et al. Safety and efficacy of chymopapain (chymodiactin) in herniated nucleus pulposus with sciatica. *JAMA* 1983; 249:2489–94.
96. Karppinen J, Malmivaara A, Kurunlahti M, et al. Periradicular infiltration for sciatica: a randomized controlled trial. *Spine* 2001;26:1059–67.
97. Klein R, Eek B, DeLong W, et al. A randomized double-blind trial of dextrose-glycerine-phenol injections for chronic, low back pain. *J Spinal Disord* 1993;6:23–33.
98. Klenerman L, Greenwood R, Davenport H, et al. Lumbar epidural injections in the treatment of sciatica. *Br J Rheum* 1984;23:35–8.
99. Kraemer J, Ludwig J, Bickert U, et al. Lumbar epidural perineural injection: a new technique. *Eur Spine J* 1997;6:357–61.
100. Leclaire R, Fortin L, Lambert R, et al. Radiofrequency facet joint denervation in the treatment of low back pain: a placebo-controlled clinical trial to assess efficacy. *Spine* 2001;26:1411–6.
101. Lilius G, Lassonen AM, Myllynen P, et al. The lumbar facet joint syndrome–significance of inappropriate signs. A randomized, placebo-controlled trial. *French J Orthop Surg*. 1989;3:479–86.
102. Mathews J, Mills S, Jenkins V, et al. Back pain and sciatica: controlled trials of manipulation, traction, sclerosant and epidural injection. *Br J Rheumatol* 1987;26:416–23.
103. Ng L, Chaudhary N, Sell P. The efficacy of corticosteroids in periradicular infiltration for chronic radicular pain. A randomized, double-blind, controlled trial. *Spine* 2005;30:857–62.
104. North RB, Kidd DH, Farrokhi F, et al. Spinal cord stimulation versus repeated lumbosacral spine surgery for chronic pain: a randomized, controlled trial. *Neurosurgery* 2005;56:98–107.
105. North RB, Kidd DH, Lee MS. A prospective, randomized study of spinal cord stimulation versus reoperation for failed back surgery syndrome: initial results. *Sterotact Funct Neurosurg* 1994;1994:267–72.
106. North RB, Kidd DH, Lee MS. Spinal cord stimulation versus reoperation for failed back surgery syndrome: a prospective study design. *Acta Neurochir Suppl* 1995;64:106–8.
107. Ongley M, Klein R, Dorman T, et al. A new approach to the treatment of chronic low back pain. *Lancet* 1987;2:143–6.
108. Pauza KJ, Howell S, Dreyfuss P, et al. A randomized, placebo-controlled trial of intradiscal electrothermal therapy for the treatment of discogenic low back pain. *Spine J* 2004;4:27–35.
109. Ridley M, Kingsley G, Gibson T, et al. Outpatient lumbar epidural corticosteroids in the management of sciatica. *Br J Rheumatol* 1988;27:295–9.
110. Riew K, Yin Y, Gilula L, et al. The effect of nerve-root injections on the need for operative treatment of lumbar radicular pain. A prospective, randomized, controlled, double-blind study. *J Bone Joint Surg* 2000;82-A:1589–93.
111. Riew KD, Park JB, Cho YS, et al. Nerve root blocks in the treatment of lumbar radicular pain. A minimum five-year follow-up. *J Bone Joint Surg Am Vol* 2006;88:1722–5.
112. Rogers P, Nash T, Schiller D, et al. Epidural steroids for sciatica. *Pain Clin* 1992;5:67–72.
113. Schwetschenau R, Ramirez A, Johnston J, et al. Double-blind evaluation of intradiscal chymopapain for herniated lumbar discs. Early results. *J Nuerosurg* 1976;45:622–7.
114. Martins A, Ramirez A, Johnston J, et al. Double blind evaluation of chemonucleolysis for herniated lumbar discs. Late results. *J Neurosurg* 1978; 49:816–27.
115. Snoek W, Weber H, Jorgensen B. Double blind evaluation of extradural methyl prednisolone for herniated lumbar discs. *Acta Orthop Scand* 1977; 48:635–41.
116. Sonne M, Chirstensen K, Hansen S, et al. Injection of steroids and local anaesthetics as therapy for low-back pain. *Scand J Rheum* 1985;14:343–5.
117. van Kleef M, Barendse G, Kessels A, et al. Randomized trial of radiofrequency lumbar facet denervation for chronic low back pain. *Spine* 1999; 24:1937–42.
118. Wilson-MacDonald J, Burt G, Griffin D, et al. Epidural steroid injection for nerve root compression. A randomised, controlled trial. *J Bone Joint Surg Br Vol* 2005;87:352–5.
119. Yelland M, Glasziou P, Bogduk N, et al. Prolotherapy injections, saline injections, and exercises for chronic low-back pain: a randomized trial. *Spine* 2004;29:9–16.
120. Zahaar M. The value of caudal epidural steroids in the treatment of lumbar neural compression syndromes. *J Neurol Orthop Med Surg* 1991;12: 181–4.
121. Abdi S, Datta S, Lucas L. Role of epidural steroids in the management of chronic spinal pain: a systematic review of effectiveness and complications: a systematic review. *Pain Physician* 2005;8:127–43.
122. Boswell M, Colson J, Spillane W. Therapeutic facet joint interventions in chronic spinal pain: a systematic review of their role in chronic spinal pain management and complications. *Pain Physician* 2005;8:101–14.
123. Boswell M, Hansen H, Trescot A, et al. Epidural steroids in the management of chronic spinal pain and radiculopathy. *Pain Physician* 2003;6: 319–34.
124. Cannon D, Aprill C. Lumbosacral epidural steroid injections. *Arch Phys Med Rehabil* 2000;81(suppl 1):S87–98.
125. Chou LH, Lew HL, Coelho PC, et al. Intradiscal electrothermal annuloplasty. *Am J Phys Med Rehabil* 2005;84:538–49.
126. Difazio M, Jabbari B. A focused review of the use of botulinum toxins for low back pain. *Clin J Pain* 2002;18(suppl 6):S155–62.
127. Gibson JN, Grant IC, Waddell G. The Cochrane review of surgery for lumbar disc prolapse and degenerative lumbar spondylosis. *Spine* 1999;24: 1820–32.
128. Gibson JN, Grant IC, Waddell G. Surgery for lumbar disc prolapse. *Cochrane Database Syst Rev* 2000:CD001350.
129. Haselkorn JK. Epidural steroid injections and the management of sciatica: a meta-analysis [Abstract]. *Arch Phys Med Rehabil* 1995;76:1037.
130. Heary RF. Intradiscal electothermal annuloplasty: The IDET procedure. *J Spinal Disord* 2001;14:353–60.
131. Hooten WM, Martin DP, Huntoon MA. Radiofrequency neurotomy for low back pain: evidence-based procedural guidelines. *Pain Med* 2005;6: 129–38.
132. Kepes ER, Duncalf D. Treatment of backache with spinal injections of local anesthetics, spinal and systemic steroids. A review. *Pain* 1985;22:33–47.
133. Kim S, Stitik T, Foye P, Greenwald B, et al. Critical review of prolotherapy for osteoarthritis, low back pain, and other musculoskeletal conditions: a physiatric perspective. *Am J Phys Med Rehabil* 2004;83:379–89.
134. Koes B. Epidural steroid injections for low back pain and sciatica: an updated systematic review of randomized clinical trials. *Pain Digest* 1999;9: 241.
135. Koes BW, Scholten RJ, Mens JM, et al. Efficacy of epidural steroid injections for low-back pain and sciatica: a systematic review of randomized clinical trials. *Pain* 1995;63:279–88.
136. Manchikanti L, Singh V, Vilims B, et al. Medial branch neurotomy in management of chronic spinal pain: systematic review of the evidence. *Pain Physician* 2002;5:405–18.
137. McKenzie-Brown AM, Shah RV, et al. A systematic review of sacroiliac joint interventions. *Pain Physician* 2005;8:115–25.
138. Nelemans PJ, de Bie RA, de Vet HCW, et al. Injection therapy for subacute and chronic benign low-back pain. *Cochrane Database Syst Rev* 1999: CD001824.
139. Rozenberg S. Glucocorticoid therapy in common lumbar spinal disorders. *Rev Rhum Engl Ed* 1998;65:649–55.
140. Rozenberg S, Dubourg G, Khalifa P, et al. Efficacy of epidural steroids in low back pain and sciatica. A critical appraisal by a French Task Force of randomized trials. Critical Analysis Group of the French Society for Rheumatology. *Rev Rhum Engl Ed* 1999;66:79–85.

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

141. Scheer S, Radack K, O'Brien D. Randomized controlled trials in industrial low back pain relating to return to work, part 2. Discogenic low back pain. *Arch Phys Med Rehabil* 1996;77:1189–97.
142. Stevens CD, Dubois RW, Larequi-Lauber T, et al. Efficacy of lumbar discectomy and percutaneous treatments for lumbar disc herniation. *Sozialund Praventivmedizin*. 1997;42:367–79.
143. Turner JA, Loeser JD, Bell KG. Spinal cord stimulation for chronic low back pain: a systematic literature synthesis. *Neurosurgery* 1995;37: 1088–95.
144. van Tulder MW, Koes BW, Bouter LM. Conservative treatment of acute and chronic nonspecific low back pain: a systematic review of randomized controlled trials of the most common interventions. *Spine* 1997;22: 2128–56.
145. Watts RW, Silagy CA. A meta-analysis on the efficacy of epidural corticosteroids in the treatment of sciatica. *Anaesth Intensive Care* 1995;23: 564–9.
146. Wetzel FT, Hassenbusch S, Oakley JC, et al. Treatment of chronic pain in failed back surgery patients with spinal cord stimulation: a review of current literature and proposal for future investigation. *Neuromodulation* 2000;3:59–74.
147. Wetzel FT, McNally TA, Phillips FM. Intradiscal electrothermal therapy used to manage chronic discogenic low back pain: new directions and interventions. *Spine* 2002;27:2621–6.
148. Yelland MJ, Del Mar C, Pirozzo S, et al. Prolotherapy injections for chronic low-back pain. *Cochrane Database Syst Rev* 2004:CD004059.
149. Bogduk N, Karasek M. Two-year follow-up of a controlled trial of intradiscal electrothermal annuloplasty for chronic low back pain resulting from internal disc disruption. *Spine J* 2002;2:343–50.
150. Dallas T, Lin R, Wu W, et al. Epidural morphine and methylprednisolone for low-back pain. *Anesthesiology* 1987;67:408–11.
151. Frost F. A controlled, double-blind comparison of mepivacaine injection versus saline injection for myofascial pain. *Lancet* 1980;1:499–501.
152. Glynn C, Dawson D, Sanders R. A double-blind comparison between epidural morphine and epidural clonidine in patients with chronic noncancer pain. *Pain* 1988;34:123–8.
153. Gobel H, Heinze A, Reichel G, et al. Dysport myofascial pain study g. Efficacy and safety of a single botulinum type A toxin complex treatment (Dysport) for the relief of upper back myofascial pain syndrome: results from a randomized double-blind placebo-controlled multicentre study. *Pain* 2006;125:82–8.
154. Kapural L, Hayek S, Malak O, et al. Intradiscal thermal annuloplasty versus intradiscal radiofrequency ablation for the treatment of discogenic pain: a prospective matched control trial. *Pain Med* 2005;6:425–31.
155. Lew HL, Lee EH, Castaneda A, et al. Therapeutic use of botulinum toxin type A in treating neck and upper-back pain of myofascial origin: a pilot study. *Arch Phys Med Rehabil* 2008;89:75–80.
156. Maugars Y, Mathis C, Berthelot JM, et al. Assessment of the efficacy of sacroiliac corticosteroid injections in spondylarthropathies: a double-blind study. *Rheumatology* 1996;35:767–70.
157. Pneumaticos SG, Chatziioannou SN, Hipp JA, et al. Low back pain: prediction of short-term outcome of facet joint injection with bone scintigraphy. *Radiology* 2006;238:693–8.
158. Vad VB, Bhat AL, Lutz GE, et al. Transforaminal epidural steroid injections in lumbosacral radiculopathy: a prospective randomized study. *Spine* 2002; 27:11–6.
159. Yates D. A comparison of the types of epidural injection commonly used in the treatment of low back pain and sciatica. *Rheum Rehabil* 1978;17: 181–6.
160. Li M, Goldberger BA, Hopkins C. Fatal case of BOTOX-related anaphylaxis? *J Forensic Sci* 2005;50:169–72.
161. Aldrete JA. Epidural injections of indomethacin for postlaminectomy syndrome: a preliminary report. *Anesth Analg*. 2003;96:463–8.
162. Buchner N, Zeifang F, Brocai D, et al. Epidural corticosteroid injection in the conservative management of sciatica. *Clin Orthop Relat Res* 2000;375: 149–56.
163. Buttermann GR. Treatment of lumbar disc herniation: epidural steroid injection compared with discectomy. A prospective, randomized study. *J Bone Joint Surg Am Vol* 2004;86-A:670–9.
164. Dashfield AK, Taylor MB, Cleaver JS, et al. Comparison of caudal steroid epidural with targeted steroid placement during spinal endoscopy for chronic sciatica: a prospective, randomized, double-blind trial. *Br J Anaesth* 2005;94:514–9.
165. Devulder J, Deene P, De Laat M, et al. Nerve root sleeve injections in patients with failed back surgery syndrome: a comparison of three solutions. *Clin J Pain* 1999;15:132–5.
166. Helsa PE, Breivik H. Epidural analgesia and epidural steroid injection for treatment of chronic low back pain and sciatica. *Tidsskr Nor Laegeforen* 1979;21:511–3.
167. Kolsi I, Delecrin J, Berthelot JM, et al. Efficacy of nerve root versus interspinous injections of glucocorticoids in the treatment of disk-related sciatica. A pilot, prospective, randomized, double-blind study. *Joint Bone Spine* 2000;67:113–8.
168. McGregor AH, Anjarwalla NK, Stambach T. Does the method of injection alter the outcome of epidural injections? *J Spinal Disord* 2001;14:507–10.
169. Meadeb J, Rozenberg S, Duquesnoy B. Forceful sacroccoccygeal injections in the treatment of postdiscectomy sciatica: a controlled study versus gluccocorticoid injections. *Joint Bone Spine* 2001;68:43–9.
170. Pirbudak L, Karakurum G, Oner U, et al. Epidural corticosteroid injection and amitriptyline for the treatment of chronic low back pain associated with radiculopathy. *Pain Clin* 2003;15:247–53.
171. Pirbudak L, Karakurum G, Satana T, et al. Epidural steroid injection and amitriptyline in the management of acute low back pain originating from lumbar disc herniation. *Artroplasti Artroskopik Cerrahi* 2003;14:89–93.
172. Revel M, Auleley G, Alaoui S. Forceful epidural injections for the treatment of lumbosciatic pain with post-operative lumbar spinal fibrosis. *Rev Rhum Engl Ed* 1996;63:270–7.
173. Rocco A, Frank E, Kaul A, et al. Epidural steroids, epidural morphine and epidural steroids combined with morphine in the treatment of postlaminectomy syndrome. *Pain* 1989;36:297–303.
174. Serrao JM, Marks RL, Morley SJ, et al. Intrathecal midazolam for the treatment of chronic mechanical low back pain: a controlled comparison with epidural steroid in a pilot study. *Pain* 1992;48:5–12.
175. Thomas E, Cyteval C, Abiad L, et al. Efficacy of transforaminal versus interspinous corticosteroid injections in discal radiculalgia–a prospective, randomized, double-blind study. *Clin Rheumatol* 2003;22:299–304.
176. Cuckler JM, Bernini PA, Wiesel SW, et al. The use of epidural steroids in the treatment of lumbar radicular pain. A prospective, randomized, double-blind study. *J Bone Joint Surg Am Vol* 1985;67:63–6.
177. Glaser SE, Falco F. Paraplegia following a thoracolumbar transforaminal epidural steroid injection. *Pain Physician* 2005;8:309–14.
178. Hooten WM, Mizerak A, Carns PE, et al. Discitis after lumbar epidural corticosteroid injection: a case report and analysis of the case report literature. *Pain Med* 2006;7:46–51.
179. Huntoon MA, Martin DP. Paralysis after transforaminal epidural injection and previous spinal surgery. *Reg Anesth Pain Med* 2004;29:494–5.
180. Fuchs S, Erbe T, Fischer HL, et al. Intraarticular hyaluronic acid versus glucocorticoid injections for nonradicular pain in the lumbar spine. *J Vasc Interv Radiol* 2005;16:1493–8.
181. Lilius G, Harilainen A, Laasonen M, et al. Chronic unilateral low back pain. Predictors of outcome of facet joint injections. *Spine* 1990;15:780–2.
182. Manchikanti L, Damron KS, Cash KA, et al. Evaluation of lumbar facet joint nerve blocks in the management of chronic low back pain: a preliminary report of a randomized, double-blind controlled trial. *Clin Trial Pain Physician*. 2007;10:425–40.
183. Manchikanti L. Effectiveness of lumbar facet joint nerve blocks in chronic low back pain: a randomized clinical trial. *Pain Physician* 2001;4:101–17.
184. Marks RC, Houston T, Thulbourne T. Facet joint injection and facet nerve block: a randomised comparison in 86 patients with chronic low back pain. *Pain* 1992;49:325–8.
185. Mayer TG, Gatchel RJ, Keeley J, et al. A randomized clinical trial of treatment for lumbar segmental rigidity. *Spine* 2004;29:2199–205.
186. Bontoux D, Alcalay M, Debiais F, et al. Treatment of lumbar disk hernia by intra-disk injection of chymopapain or triamcinolone hexacetonide. Comparative study of 80 cases. *Rev Rhum Mal Osteoartic* 1990;57:327–31.
187. Bourgeois P, Benoist M, Palazzo E, et al. Multicenter randomized double-blind study of triamcinolone hexacetonide versus chymopapain in the treatment of disc lumbosciatica. Initial results at 6 months. *Rev Rhum Mal Osteoartic* 1988;55:767–9.
188. Graham C. Chemonucleolysis: a double blind study comparing chemonucleolysis with intra discal hydrocortisone: in the treatment of backache and sciatica. *Clin Orthop* 1976:179–92.
189. Benoist M, Bonneville JF, Lassale B, et al. A randomized, double-blind study to compare low-dose with standard-dose chymopapain in the treatment of herniated lumbar intervertebral discs. *Spine* 1993;18:28–34.
190. Burton, Waddell GT. Information and advice to patients with back pain does have a positive effect: an RCT of an educational booklet in primary care [Abstract]. *J Bone Joint Surg Br Vol* 1999;82(suppl 1):3.
191. Crawshaw C, Frazer AM, Merriam WF, et al. A comparison of surgery and chemonucleolysis in the treatment of sciatica. A prospective randomized trial. *Spine* 1984;9:195–8.
192. Ejeskar A, Nachemson A, Herberts P, et al. Surgery versus chemonucleolysis for herniated lumbar discs. A prospective study with random assignment. *Clin Orthop* 1983;174:236–42.

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

193. Feldman J, Menkes CJ, Pallardy G, et al. Double-blind study of the treatment of disc lumbosciatica by chemonucleolysis. *Rev Rhum Mal Osteoartic* 1986;53:147–52.
194. Hedtmann A, Fett H, Steffen R, et al. Chemonucleolysis using chymopapain and collagenase. 3-year results of a prospective randomized study. *Z Orthop Ihren Grenzgeb* 1992;130:36–44.
195. Krugluger J, Knahr K. Chemonucleolysis and automated percutaneous discectomy–a prospective randomized comparison. *Int Orthop* 2000;24:167–9.
196. Lavignolle B, Vital J, Baulny D, et al. Comparative study of surgery and chemonucleolysis in the treatment of sciatica caused by a herniated disc. *Acta Orthop Belg* 1987;53:244–9.
197. Muralikuttan KP, Hamilton A, Kernohan WG, et al. A prospective randomized trial of chemonucleolysis and conventional disc surgery in single level lumbar disc herniation. *Spine* 1992;17:381–7.
198. Revel M, Payan C, Vallee C, et al. Automated percutaneous lumbar discectomy versus chemonucleolysis in the treatment of sciatica. A randomized multicenter trial. *Spine* 1993;18:1–7.
199. Steffen R, Luetke A, Wittenbert R, et al. A prospective comparative study of chemonucleolysis and laser discectomy. *Orthop Trans* 1996;20:388.
200. Steffen R, Wittenberg R, Kraemer J. Chemonucleolysis versus laser disc decompression–a prospective randomised trial. *J Bone Joint Surg* 1997;79-B:247.
201. van Alphen HA, Braakman R, Bezemer PD, et al. Chemonucleolysis or discectomy: a randomized multicenter trial. *J Neurosurg* 1989;70:869–75.
202. Yu L, Wu N, Zhang K. An investigation on the treatment of intraprotrusion or intradisk injection with collagenase for lumbar disc herniation. *Chin J Clin Rehabil* 2002;6:598–9.
203. Day PL. Early, interim, and long-term observations on chemonucleolysis in 876 patients with special comments on the lateral approach. *Clin Orthop* 1974;99:64.
204. Wiltse LL, Widell EH, Yuan HA. Chymopapain chemonucleolysis in lumbar disk disease. *JAMA* 1975;231:474–9.
205. Agre K, Wilson RR, Brim M, et al. Chymodiactin post marketing surveillance: demographic and adverse experience data. *Spine* 1984;9:479–86.
206. Brown MD, Tompkins JS. Chemonucleolysis (discolysis) with collagenase. *Spine* 1989;11:123–30.
207. Sanders M, Zuurmond WW. Percutaneous intraarticular lumbar facet joint denervation in the treatment of low back pain: a comparison with percutaneous extra-articular lumbar facet denervation. *Pain Clin* 1999;11:329–35.
208. Bogduk N. Lumbar radiofrequency neurotomy. *Clin J Pain* 2006;22:409.
209. Gofeld M. Radiofrequency facet denervation: a randomized control placebo versus sham procedure. *Clin J Pain* 2006;22:410–1.
210. Gerszten PC, Welch WC, McGrath PM, et al. A prospective outcomes study of patients undergoing intradiscal electrothermy (IDET) for chronic low back pain. *Pain Physician* 2002;5:360–4.
211. Ercelen O, Bulutcu E, Oktenoglu T, et al. Radiofrequency lesioning using two different time modalities for the treatment of lumbar discogenic pain: a randomized trial. *Spine* 2003;28:1922–7.
212. Niemisto L, Lahtinen-Suopanki T, Rissanen P, et al. A randomized trial of combined manipulation, stabilizing exercises, and physician consultation compared to physician consultation alone for chronic low back pain. *Spine* 2003;28:2185–91.
213. Evans W. Intrasacral epidural infections in the treatment of sciatica. *Lancet* 1930;2:1225–9.
214. Tachihara H, Sekiguchi M, Kikuchi S, et al. Do corticosteroids produce additional benefit in nerve root infiltration for lumbar disc herniation? *Spine* 2008;33:743–7.
215. Ioannidis JP, Evans SJ, Gotzsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med*. 2004;141:781–8.
216. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical vs nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT): a randomized trial. *JAMA* 2006;296:2441–50.

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

Health Technology Clinical Committee
Findings and Coverage Decision
Topic: **Spinal Injections**
Meeting Date: **March 18th, 2011**
Final Adoption: **June 17th, 2011**

Number and Coverage Topic

20110318B – Spinal Injections

HTCC Coverage Determination

Therapeutic Medial Branch Nerve Block injections, Intradiscal injections and Facet injections are **not a covered benefit**

Therapeutic Lumbar Epidural Injections; Cervical-thoracic Epidural Injections and Sacroiliac Joint Injections are a **covered benefit** for the treatment of chronic pain

HTCC Reimbursement Determination

- **Limitations of Coverage**
  - Therapeutic Epidural Injections in the lumbar or cervical-thoracic spine for chronic pain are a covered benefit when all of the following conditions are met:
    - For treatment of radicular pain
    - With fluoroscopic guidance or CT guidance
    - After failure of conservative therapy
    - No more than two without clinically meaningful improvement in pain and function, and
    - Maximum of 3 in 6 months
  - Therapeutic Sacroiliac Joint Injections for chronic pain is a covered benefit when all of the following conditions are met:
    - With Fluoroscopic guidance or CT guidance
    - After failure of conservative therapy, and
    - No more than one without clinically meaningful improvement in pain and function, subject to agency review

- **Non-Covered Indicators**
  - Therapeutic Medial Branch Nerve Block injections; Intradiscal injections and Facet injections are not a covered benefit.

- **Agency Contact Information**

Agency	Contact Phone Number
Labor and Industries	1-800-547-8367
Public Employees Health Plan	1-800-762-6004
Health and Recovery Services Administration	1-800-562-3022

Health Technology Background

The Spinal Injections topic was selected and published in December 2009 to undergo an evidence review process. The evidence based technology assessment report indicates that an estimated 75% of the population has had an episode of back pain at some point in their life. While most acute back pain resolves within a few months, surveys report that approximately 5% of the population has chronic back pain, a percentage which implicates significant social and economic impacts. The risk of spinal pain increases with age as a result of disc disease and spinal degeneration. Those affected can have disabling symptoms that can dramatically affect their quality of life and ability to perform a variety of activities. Chronic spinal pain can be attributed to a number of pathologies, including (but not limited to) degenerative disc disease (DDD), herniated nucleus pulposus (HNP) (or herniated/slipped disc), spinal stenosis, radiculopathy, failed back surgery syndrome (FBSS), facet joint syndrome, and whiplash.

Treatment for chronic back pain typically begins with the identification of the underlying cause of pain and follows with conventional medical management (CMM), which varies with the diagnosis. CMM may include conservative/ non-invasive interventions such as physical therapy and rehabilitation, pharmaceutical pain management, psychological therapy and coping skills, exercise, education, antidepressants, cognitive behavioral therapy and supported self-management, spinal manipulation, electrical stimulation, injections outside the spine, implanted devices, acupuncture/acupressure, and modified work.

Patients who don't respond to non-invasive treatment are typically referred for more invasive and non-surgical therapies such as spinal injections in an attempt to provide pain relief. Spinal injections involve the injection of an anti-inflammatory agent such as a steroid and/or an anesthetic into the spine or space around the spinal nerves and joints. One of the theoretical advantages of spinal injections is that they deliver the treatment medication directly to the site involved in the source of pain. Types of spinal injection include epidural, facet joint, intradiscal, and sacroiliac joint injections. Spinal injections can be used for diagnostic and therapeutic purposes.

In November 2010, the HTA posted a draft and then followed with a final report from a contracted research organization that reviewed publicly submitted information; searched, summarized, and evaluated trials, articles, and other evidence about the topic. The comprehensive, public and peer reviewed Spinal Injections report is 299 pages, and identified a relatively large amount of literature.

An independent group of eleven clinicians who practice medicine locally meet in public to decide whether state agencies should pay for the health technology based on whether the evidence report and other presented information shows it is safe, effective and has value. The committee met on March 18th, reviewed the report, including peer and public feedback, and heard public and agency comments. Meeting minutes detailing the discussion are available through the HTA program or online at http://www.hta.hca.wa.gov under the committee section.

Committee Findings

Having considered the evidence based technology assessment report and the written and oral comments, the committee identified the following key factors and health outcomes, and evidence related to those health outcomes and key factors:

1. Evidence availability and technology features

The committee concludes that the best available evidence on Spinal Injections has been collected and summarized. The evidence is presented below:

- The evidence based technology assessment report estimates 75% of the population has an episode of back pain at some point in their life. While most acute back pain resolves within a few months, surveys report that approximately 5% of the population has chronic back pain, with significant social and economic impacts. Those affected can have disabling symptoms that can dramatically affect their quality of life and ability to perform a variety of activities. The source and pathology of chronic spinal pain is not well understood but has been attributed degenerative disc disease (DDD), herniated nucleus pulposus (HNP) (or herniated/slipped disc), spinal stenosis, radiculopathy, failed back surgery syndrome (FBSS), facet joint syndrome, among other causes.
- The evidence based technology assessment report indicates treatment for chronic back pain typically begins with the identification (or ruling out) of underlying cause of pain and beginning conventional medical management (CMM). CMM may include conservative/ non-invasive interventions such as physical therapy and rehabilitation, pharmaceutical pain management, psychological therapy and coping skills, exercise, education, antidepressants, cognitive behavioral therapy and supported self-management, spinal manipulation, electrical stimulation, injections outside the spine, implanted devices, acupuncture/acupressure, and modified work.
- The evidence based technology assessment report indicates that a small percentage of non-responsive patients may proceed to invasive therapies, including spinal injections. Spinal injections are not curative but are intended to provide pain relief and functional improvement for up to several months. Spinal injections involve the injection of an anti-inflammatory agent such as a steroid and/or an anesthetic into the spine or space around the spinal nerves and joints. One of the theoretical advantages of spinal injections is that they deliver medication directly to the site thought to be the source of pain. Types of spinal injection include epidural, facet joint, intradiscal, and sacroiliac joint injections. Spinal injections can be used for diagnostic and therapeutic purposes. According to one study examining Medicare claims of lumbosacral injections, the number of epidural steroidal injections increased 271% and the number of facet injections increased 231% from 1994 to 2001. A similar study found that lumbar facet joint injections/diagnostic blocks increased 161% from 2002 to 2006.
- Despite dramatic growth in procedures, evidence about the impact of spinal injections on important patient oriented outcomes related to impact on pain, physical function, opioid use; return to work; quality of life; patient satisfaction; avoidance of more invasive surgery; expected duration of impact; need for repeat procedures; frequency and type of harms; as well as clinical impacts of multilevel or procedure differences and any evidence about differential effect based on different patient, social or provider characteristics; different injection types; and impact of cost is needed.
- The evidence based technology assessment report indicates that the Spinal injection evidence base is extensive: initial search resulted in over 2,700 potential citations; and based on evaluation against inclusion criteria, 1 Systematic review; 22 RCTs, 24 Observational Studies and two economic studies were included.
    - Evidence was identified on five injection types: epidural (lumbar and cervical); facet joint; sacroiliac; intradiscal injections and medial branch blocks.
    - Key strengths of the overall body of evidence are a large evidence base including randomized clinical trials.

- Limitations in the overall body of evidence: despite well validated measures to evaluate treatment outcomes, evidence is limited by the variety of different measures or non-validated measures used; most studies were limited by a focus on one outcome - impact on short term pain; studies not including a placebo arm are limited when measuring subjective improvement in pain; many studies were limited by short duration (3 month or less) for treatment of a chronic condition; there remains uncertainty over clinically meaningful improvement for pain and function; and the variety of injection methods and types.

2. Is the technology safe?

The committee discussed multiple key factors and health outcomes that were important for consideration in their overall decision on whether the technology is safe. Summary of committee considerations follows.

- Major Complications: the evidence based technology assessment report indicated that major reported complications of spinal injection include dural puncture; subarachnoid puncture and angina pectoris, though rates are rare.
  - There were no cases of death or paralysis related to the procedure in the included studies, though death unrelated to the procedure was reported in 10 of 1146 patients in the RCTs, and there have been case reports of death and paralysis in the published literature.
  - For dural or subarachnoid punctures, or other life threatening complications, the reported rates ranged from 3 in 710 injections to 5 in 7240 (cervical) and 1 in 1556 injections to 1 in 10,416 injections for lumbar.
  - Vascular Puncture: the evidence based technology assessment report indicated the mean incidence of intravascular puncture following fluoroscopically guided lumbar spinal injections was 10.18% (range, 1.9–22%) as reported in five case series designed to assess its incidence.
- Minor Complications: the evidence based technology assessment report indicated that minor complications are more common but are generally transient in nature. The overall minor complication rate ranged from 0.06% to 16.3% of injections or patients in 19 RCTs and 14 case series, and complications included: pain at the injection site, increased radicular pain/numbness/weakness, nerve root irritation, superficial infections, sympathetic blockade, facial flushing, vasovagal reactions/fainting, headache, gastric complaints, dizziness, pruritis, irregular periods, and insomnia.
- Radiation Exposure to the Physician: the evidence based technology assessment report indicated the with proper protective measures, total radiation exposure was within normal limits following a mean of 923 procedures (range, 100 – 1819) with an average length of radiation exposure of 9.8 seconds/procedure (range, 4.9 – 15.2) in all five case series we identified.
  - The evidence based technology assessment report reported that approximately 50% of four million interventional medical procedures per year are performed under fluoroscopic guidance. Fluoroscopy for spinal injections is routinely used to ensure correct needle placement, accurate delivery of the injectate, and avoidance of complications. Incorrect needle placement during spinal injections without the use of fluoroscopy has been reported by various studies in 12.5% to 38.3% of patients. A C-arm fluoroscope allows the X-ray tube to be moved around the prone patient and an image intensifier enhances the image, making it easier to interpret. Although studies have shown that radiation exposure to physicians using fluoroscopy for spinal injections is within safety limits, other methods, including ultrasound and CT, are being investigated as non-radioactive or lower radioactive methods of needle guidance.

3. Is the technology effective?

The committee discussed multiple key factors and health outcomes that were important for consideration in their overall decision on whether the technology is effective. Summary of committee considerations follows.

- Discussion focused on the following categories of injections: lumbar epidural; cervical/thoracic epidural; facet joint injection; sacroiliac joint injection; medial branch block; and intradiscal injection. Further differentiation was not focused on as the evidence based technology report indicated low to very low overall strength of evidence of different impact. The low level of evidence reported no consistent differential impact based on the approach to administering the injection; the diagnosis, pre-injection pain intensity; type of steroid, gender, age or other patient characteristics.
- **Epidural Steroid Injections for lumbar or low back pain with sciatica or radiculopathy** was highly studied and reported on; however, the overall strength of evidence is low based on the individual trial limitations and the inconsistency in results. Low back pain with sciatica or radiculopathy *the evidence is mixed* about the impact of spinal injection on pain (and in some studies function); with some studies showing a inferior results compared to placebo or other interventions and some studies showing a positive result.
  - When compared to placebo for caudal or interlaminar: *In the short-term* (≤ 3 months) there was mixed evidence based on data from twenty RCTs, seventeen of which were included in the Chou/APS SR (seven were considered to be higher-quality trials). Seven of seventeen studies included in the SR reported no benefit or inferior results while another seven reported positive results and three reported unclear results. Three LoE IIb RCTs published after the SR were added here, two reported on pain (both negative) and three on function (two negative and one positive) at three months. *In the long-term* (> 3 months) there was mixed evidence based on data from twelve RCTs, nine of which were included in the Chou/APS SR. Seven of nine studies included in the SR reported no benefit or inferior results while positive results were reported by one study and another reported mixed results. Regarding the more recent RCTs included here, two reported on pain (both negative at twelve months, although one was positive at six months) and three on function (mixed results, one positive, one mixed, and one negative). (SoE = Low)
  - When compared to placebo for transforaminal: mixed evidence based on data from four RCTs, two of which were included in the Chou/APS SR and considered to be higher-quality and two of which were more recent LoE IIb studies. In terms of pain relief, the data suggest a benefit at two weeks (one study), mixed results at one month (two studies- one positive and one negative), and no benefit by 3 months. No benefit in function was reported at three months by two studies. Long-term data were mixed as reported by two higher-quality RCTs, both of which were reported in the Chou/APS SR, with one study reported positive results while the other showed no benefit. When compared to intramuscular injections, transforaminal steroid injections were superior to intramuscular injections in terms of pain relief at one month based on data from one LoE IIb RCT. (SoE = Low)
- **Epidural Steroid Injections for lumbar or low back pain without sciatica or radiculopathy** was also studied and reported on, and the overall strength of evidence is low to moderate based on the individual trial limitations and indication studied. The evidence indicates *no benefit* of spinal injections compared either to placebo, physical therapy, trigger point injection, discectomy or dry needling.
  - Low back pain (without sciatica or radiculopathy) compared to placebo showed no benefit based on data from three RCTs, one of which was included in the Chou/APS SR

and considered to be a lower-quality trial. The two more recent RCTs rated IIb also reported no benefit in pain, function, or opioid use at three months or in employment at twelve months. (SoE = Moderate)

- Spinal Stenosis compared to placebo: *In the short-term* (24 hours – 3 months), there was no benefit based on data from four RCTs, three of which was included in the Chou/APS SR; one was considered to be a higher-quality trial. Three of four studies reported no benefit; one study reported improved walking distance at one week. In a recent RCT, LoE IIb there was no benefit in pain, function, or opioid use at three months. (SoE = moderate). *In the long-term* (13 – 30 months), there was no benefit based on data from two RCTs as reported in the Chou/APS SR. (SoE = Low)
- Failed back surgery syndrome compared to placebo: no benefit based on data from three RCTs, two of which were included in the Chou/APS SR and considered to be lower-quality trials. In the one recent LoE IIb RCT, there was no benefit in pain, function, or opioid use at three months. (SoE = Moderate)
- Spinal Stenosis compared to physical therapy or control: no benefit in terms of pain, function, or quality of life at three and six months based on data from one LoE IIb RCT. (SoE = Very Low)

- **Epidural Steroid Injections for cervical pain** reported overall strength of evidence of very low based on small number of trials, trial limitation and inconsistent results. The evidence indicates *mixed benefit* of epidural cervical spinal injections.
  - For neck pain <u>with</u> disc herniation and radiculitis (comparator = placebo): *no benefit* in terms of pain, function, or opioid use at both three and twelve months or on employment at twelve months based on data from one LoE IIb RCT. (SoE = Very Low)
  - Neck pain <u>without</u> disc herniation and radiculitis (comparator = placebo): *no benefit* in terms of pain, function, or opioid use at both three and twelve months or on employment at twelve months based on data from one LoE IIb RCT. (SoE = Very Low)
  - Neck pain with disc compression and radiculitis (comparator = intramuscular injection): epidural injections were *superior* to intramuscular injections in the posterior neck in terms of pain, analgesic use, and employment at one week and twelve months based on data from one LoE IIb RCT. (SoE = Very Low)
- Facet Joint Steroid Injections overall had low strength of evidence of no benefit based on four RCTs.
  - Confirmed or presumed lumbar facet joint pain compared to placebo: no benefit in the first three months based on data from two RCTs included in the Chou/APS SR, one of which was considered to be lower-quality. Although one of the studies reported a statistically meaningful benefit at six months in patient improvement following steroid injection, the rationale for this late response is not clear. (SoE = Low)
  - Non-radicular back pain and facet joint osteoarthritis compared to hyaluronic acid: no benefit in the injection of steroids versus hyaluronic acid into the facet joint at six months based on data from one higher-quality RCT included in the Chou/APS SR. (SoE = Low)
  - Confirmed cervical facet joint pain compared to placebo: no benefit in terms of the length of pain relief based on data from one LoE IIb RCT. No long-term data was reported. (SoE = Very Low)
- Sacroiliac Joint Steroid Injections had low overall strength of evidence of benefit based on one RCT.
  - For sacroiliac Joint Pain, compared to placebo: sacroiliac joint injections were superior to placebo injections based on data from one higher-quality RCT included in the Chou/APS SR. (SoE = Low)

- Intradiscal Injections overall had moderate strength of evidence of no benefit based on seven RCTs.
  - For discogenic back pain, steroid injection compared to placebo: no benefit based on data from three RCTs included in the Chou/APS SR, one of which was higher-quality. (SoE = Moderate)
  - For sciatica compared to chemotherapy: no benefit based on data from three RCTs included in the Chou/APS SR, one of which was higher-quality. (SoE = Moderate)
  - For low back pain without radiculopathy using neurolytic agent compared to placebo: intradiscal injections with methylene blue were superior to placebo injections in terms of pain, function, patient satisfaction, and analgesic use in the long-term (6-24 months) based on data from one LoE IIa RCT. (SoE = Low)
- Medial Branch Blocks overall had low to very low strength of evidence of no benefit based on four RCTs.
  - For confirmed lumbar facet joint pain compared to placebo: no benefit in terms of pain or function at both three and twelve months or on opioid use at twelve months based on data from one LoE IIb RCT. (SoE = Very Low)
  - For presumed lumbar facet joint pain compared to Sarapin: no benefit in injections with Sarapin with or without steroid based on data from one higher-quality and one lower-quality RCT included in the Chou/APS SR. (SoE = Low)
- For confirmed cervical facet joint pain compared to placebo: no benefit in terms of pain or function at both three and twelve months or on opioid use or employment at twelve months based on data from one LoE IIb RCT. (SoE = Very Low)

4. Special Populations?

- Approach of the Epidural Steroid Injection: the evidence based technology assessment report indicated no consistent evidence from a systematic review of six RCTs and two additional RCTs published since the systematic review that one approach is more efficacious in administering lumbar epidural steroid. The results of one lower quality RCT suggest that interlaminar injections may not be as efficacious as transforaminal in patients with axial only pain from spinal stenosis. However, more study is needed to verify these findings.
- Diagnosis: the evidence based technology assessment report indicated no consistent evidence that epidural steroid injections have differential efficacy or effectiveness among various diagnoses of the lumbar or cervical spine.
- Pre-injection pain intensity or duration, type of steroid, sex, age, or MRI findings: the evidence based technology assessment report indicated no consistent evidence that pre-injection pain intensity or duration, type of steroid used as injectate, sex, age or pre-injection MRI findings are associated with outcome in patients receiving epidural steroid injections of the lumbar or cervical spine.

5. Is the technology cost-effective?

The committee discussed multiple key factors that were important for consideration in their overall decision on whether the technology has value and is cost-effective. Summary of committee considerations follows.

- The evidence based technology assessment report reported no evidence that epidural steroid injections are cost effective based on data from two economic analyses. One moderately well conducted cost utility analysis (QHES 78/100) suggested that one epidural steroid injection is a more cost effective patient management strategy than up to three injections and that cost effectiveness ratios for epidural steroid injections are too high to be considered cost effective by

UK conventions. Further, the budget impact of epidural spinal injections is likely large because of high use. Poor economic data (QHES 49/100) from a second trial (Karppinen) suggested that over one year epidural steroid injections do not show cost or outcome advantages compared to saline injections, and that contained herniations may be more responsive to steroid injection than bulges or extrusions.

- The evidence based technology assessment report reported no economic data were available for facet injections, medial branch blocks, sacroiliac joint injections, or intradiscal injections or for any type of cervical injection.
    - Washington state agency utilization and cost information indicated costs for Spinal Injections of $55M for the past four years with a rising trend.

6. Medicare Decision and Expert Treatment Guidelines

Committee reviewed and discussed the Medicare Decision and expert guidelines as identified and reported in the technology assessment report.

- The Centers for Medicare and Medicaid Services have no published National coverage determinations (NCD) for any spinal injections.
- Guidelines – a search of the core sources and relevant specialty groups identified fourteen guidelines.
    - American Pain Society (APS), 2009: For patients with nonradicular low back pain, the APS is unable to assess the benefit of epidural steroid injection, facet joint steroid injection, medial branch block, or sacroiliac joint injection based on insufficient or poor evidence. Corticosteroid facet joint injection is not recommended based on moderate evidence. Intradiscal steroid injection is not recommended for treatment of nonradicular low back pain based on good evidence. For patients with radicular low back pain, the APS found moderate evidence for short-term (through three months) benefit from epidural steroid injections based on fair evidence. A recommendation for epidural steroid injection for patients with symptomatic spinal stenosis is not offered based on insufficient or poor evidence.
    - American Society of Interventional Pain Physicians, 2009: The recommendation for caudal epidural steroid injection in managing lumbar spinal pain with disc herniation and radiculitis or discogenic pain without disc herniation or radiculitis is 1A or 1B, indicating a strong recommendation where the benefits outweigh the risks of treatment. In addition, the recommendation for caudal epidural steroid injection for patients with post-lumbar laminectomy syndrome and spinal stenosis is 1B or 1C, also indicating a strong recommendation. The recommendation for use of cervical interlaminar epidural injection for disc herniation and radiculitis to achieve short-term relief is 1C. For patients seeking long-term relief, the recommendation is 2B (weak recommendation), indicating benefits are balanced with risks and burdens of treatment. In patients with spinal stenosis and discogenic pain without disc herniation and radiculitis the recommendation is 2C (very weak, with uncertainty in estimates of benefits, risk, and burden of treatment). The recommendation for lumbar transforaminal epidural injections is 1C. Intraarticular facet joint injections are not recommended. Cervical, thoracic, and lumbar facet joint nerve blocks are recommended to provide both short-term and long-term relief in the treatment of chronic facet joint pain (recommendation 1B or 1C).
    - Institute for Clinical Systems Improvement (ICSI), 2009: Epidural steroid injections and facet joint injections are classified as level I (standard, first-line) therapeutic procedures, and are recommended as part of a comprehensive treatment plan that includes pharmacologic, rehabilitative, and psychological interventions. Evidence is limited when such procedures are used alone.

- American College of Occupational and Environmental Medicine (ACOEM), 2008: Epidural glucocorticosteroid injection is recommended as a treatment option for subacute radicular pain syndromes, and as an option for second-line treatment of acute flare-ups of spinal stenosis associated with true radicular or radiculomyelopathic symptoms based on low potential harm to the patient and low costs (Evidence Rating I: insufficient evidence). Epidural glucocorticosteroid injection is not recommended to treat chronic neck pain or for dorsal spine symptoms that predominate over leg pain based on evidence that harms and cost exceed benefits to the patient (Evidence Rating C: limited evidence). The ACOEM makes no recommendation regarding the use of facet joint injection for flare-ups of neuropathic pain or chronic low back pain (Evidence Rating I: insufficient evidence). Facet joint injection is not recommended for any radicular pain syndrome, chronic non-specific axial pain, and repeat injections are not recommended for patients who failed to achieve lasting functional improvements after a prior injection for neuropathic or chronic low back pain based on evidence that treatment is ineffective or that costs or harms outweigh benefits to the patient (Evidence Rating B: moderate evidence).
- Institute for Clinical Systems Improvement (ICSI), 2008: ICSI recommends epidural steroid injection only after conservative treatment has failed and to avoid surgical intervention. ICSI finds limited evidence for the efficacy of epidural steroid injection, but indicates it may allow patients to progress with conservative treatments. Epidural steroid injection should be performed under fluoroscopy with contrast in order to prevent treatment failure.
- Work Loss Data Institute, Low back – lumbar & thoracic (acute & chronic), 2008: Epidural steroid injection and sacroiliac joint injections are recommended as part of a comprehensive treatment plan for low back pain. Specifically, epidural steroid injection is recommended to avoid surgery for severe cases with radiculopathy, but does not offer long-term functional benefit. "Series of three" epidural steroid injections, facet joint injection (multiple series, thoracic, and medical branch blocks), and intradiscal steroid injection were considered but are not recommended.
- Work Loss Data, Neck and upper back (acute & chronic), 2008: Epidural steroid injection is recommended as part of a comprehensive treatment plan for radicular pain. Specifically, epidural steroid injection is recommended to avoid surgery in severe cases with neurologic findings. Facet joint injection was considered but is not recommended.
- Work Loss Data, Pain (chronic), 2008: Epidural steroid injection is recommended as part of a comprehensive treatment plan. Facet blocks are classified as under study by the Institute and are not currently recommended.
- American Academy of Neurology, 2007: The American Academy of Neurology indicates the use of epidural steroid injections may result in a small magnitude of improvement in radicular lumbosacral pain when evaluated 2-6 weeks post-injection, but the recommendation is classified as a level C (possibly effective) due the small number of relevant studies, highly select patient population, and variation in comparison treatments in the evidence base. Epidural steroid injections are not recommended for radicular lumbosacral pain due to a lack of evidence for improvement of function, need for surgery or long-term pain relief beyond 3 months. This recommendation is classified as level B (probably ineffective based on Class I-III evidence). There was insufficient evidence to make a recommendation regarding the use of epidural steroid injections to treat cervical radicular pain.
- American College of Occupational and Environmental Medicine, 2007: The use of epidural glucocorticosteroid injection is recommended as a second-line treatment of acute spinal stenosis flare-ups, and as a treatment option for acute or subacute radicular

pain syndromes lasting at least 3 weeks after treatment with NSAIDs and when pain is not trending towards spontaneous resolution. Both treatments are recommended based on low potential harm to the patient and low costs (Evidence Rating I: insufficient evidence). The use of facet joint injections is not recommended for acute, subacute, chronic low back pain, and radicular pain syndrome based on evidence that the treatment is ineffective or that harms and cost exceed benefits to the patient (Evidence Rating B: moderate evidence). Sacroiliac joint corticosteroid injection is recommended as an option for patients with specified known cause of sacroilitis (Evidence Rating C: limited evidence). The use of epidural glucocorticosteroid injection is not recommended for acute, subacute, or chronic low back pain in the absence of radicular signs and symptoms (Evidence Rating C: limited evidence).

- American College of Physicians and the American Pain Society, 2007: Epidural steroid injection is an option for patients with prolapsed lumbar disc with persistent radicular symptoms who have not responded to noninvasive therapy. No specific recommendation is given for this or any other injection therapy of interest.
- North American Spine Society (NASS), 2007: The NASS recommends nonfluoroscopically-guided interlaminar epidural steroid injection as a treatment option for short-term symptom relief in patients with neurogenic claudication or radiculopathy. A single radiographically-guided transforaminal injection may also provide short-term symptom relief for patients with radiculopathy (Grade B: fair evidence). A multiple injection regimen of radiographically-guided transforaminal epidural steroid injection or caudal injections may provide long-term symptom relief in patients with radiculopathy or neurogenic intermittent claudication, but evidence supporting this recommendation is of poor quality.
- EuroCOST: European evidence-based guideline COST B13 Working Group on Guidelines for Chronic Low Back Pain, 2006: Epidural steroid injection, facet joint injection, and facet nerve blocks are not recommended based on a lack of evidence or conflicting evidence. Intradiscal injections are not recommended for the treatment chronic nonspecific low back pain based on evidence they are not effective (level B: moderate evidence).
- American Association of Neurological Surgeons; Congress of Neurological Surgeons, 2005: Lumbar epidural injections and facet injections are recommended as treatment options for temporary, symptomatic relief in some patients with chronic low back pain, but epidural injections are not recommended for long-term relief of pain, based on Class III evidence (unclear clinical certainty). Facet injections are not recommended as long-term treatment for low back pain based on Class I evidence (high clinical certainty).

Committee Decision

Based on the deliberations of key health outcomes, the committee decided that it had the most complete information: a comprehensive and current evidence report, public comments, and agency and state utilization information.

- The committee concluded that the current evidence on Spinal Injections demonstrates that there is sufficient evidence to cover with conditions the use of therapeutic Epidural injections in the lumbar or cervical-thoracic spine for chronic pain.
- The committee concluded that the current evidence on Spinal Injections demonstrates that there is sufficient evidence to cover with conditions therapeutic Sacroiliac joint injections for chronic pain.

- The committee concluded that the current evidence on Spinal Injections demonstrates that there is insufficient evidence to cover the other therapeutic spinal injections: Facet joint injections; medial branch block injections; and Intradiscal injections.

The committee considered all the evidence and gave greatest weight to the evidence it determined, based on objective factors, to be the most valid and reliable. Based on these findings, the committee voted to cover with conditions lumbar epidural injections. Based on these findings, the committee voted to cover with conditions cervical-thoracic epidural injections. Based on these findings, the committee voted to not cover medial branch blocks. Based on these findings, the committee voted to not cover Intradiscal injections. Based on these findings, the committee voted to not cover facet injections. Based on these findings, the committee voted to cover with conditions Sacroiliac joint injections.

Based on the evidence about the technologies' safety, efficacy, and cost-effectiveness, therapeutic Epidural Injections in the lumbar or cervical-thoracic spine is a covered benefit when all of the following conditions are met:

- For treatment of radicular pain
- With fluoroscopic guidance or CT guidance
- After failure of conservative therapy
- No more than two without clinically meaningful improvement in pain and function
- Maximum of 3 in 6 months

Based on the evidence about the technologies' safety, efficacy, and cost-effectiveness, therapeutic Sacroiliac Joint Injections for chronic pain is a covered benefit when all of the following conditions are met:

- With Fluoroscopic guidance or CT guidance
- After failure of conservative therapy
- No more than one without clinically meaningful improvement in pain and function, under agency review

Health Technology Clinical Committee Authority

Washington State's legislature believes it is important to use a scientific based, clinician centered approach for difficult and important health care benefit decisions. Pursuant to chapter 70.14 RCW, the legislature has directed the Washington State Health Care Authority, through its Health Technology Assessment program to engage in a process for evaluation process that gathers and assesses the quality of the latest medical evidence using a scientific research company and takes public input at all stages. Pursuant to RCW 70.14.110 a Health Technology Clinical Committee (HTCC) composed of eleven independent health care professionals reviews all the information and renders a decision at an open public meeting. The Washington State Health Technology Clinical Committee (HTCC) determines how selected health technologies are covered by several state agencies (RCW 70.14.080-140). These technologies may include medical or surgical devices and procedures, medical equipment, and diagnostic tests. HTCC bases their decisions on evidence of the technology's safety, efficacy, and cost effectiveness. Participating state agencies are required to comply with the decisions of the HTCC. HTCC decisions may be re-reviewed at the determination of the HCA Administrator.

Draft Coverage Guidance Review

HEALTH EVIDENCE REVIEW COMMISSION (HERC)

DRAFT COVERAGE GUIDANCE: CORONARY ARTERY CALCIUM SCORING

DATE: XX/XX/XXXX

HERC COVERAGE GUIDANCE

Coronary artery calcium scoring (CACS) should not be covered.

RATIONALE FOR GUIDANCE DEVELOPMENT

The HERC selects topics for guideline development or technology assessment based on the following principles:

- Represents a significant burden of disease
- Represents important uncertainty with regard to efficacy or harms
- Represents important variation or controversy in clinical care
- Represents high costs, significant economic impact
- Topic is of high public interest

Coverage guidance development follows to translate the evidence review to a policy decision. Coverage guidance may be based on an evidence-based guideline developed by the Evidence-based Guideline Subcommittee or a health technology assessment developed by the Heath Technology Assessment Subcommittee. In addition, coverage guidance may utilize an existing evidence report produced by one of HERC's trusted sources, generally within the last three years.

EVIDENCE SOURCES

Hayes, Inc. (2012). *Coronary artery calcium scoring to assess the risk of coronary artery disease in asymptomatic adults.* Lansdale, PA: Hayes, Inc.

National Institute for Health and Clinical Excellence (NICE). (2010). *Chest pain of recent onset: Assessment and diagnosis of recent onset chest pain or discomfort of suspected cardiac origin.* London: NICE. Retrieved August 31, 2012, from http://www.nice.org.uk/nicemedia/live/12947/47938/47938.pdf

DRAFT

U.S. Preventive Services Task Force. (2009). Using Nontraditional Risk Factors in Coronary Heart Disease Risk Assessment 2009. Retrieved August 31, 2012, from http://www.uspreventiveservicestaskforce.org/uspstf09/riskcoronaryhd/coronaryhdrs.htm

Washington State Health Care Authority Health Technology Assessment Program. (2009). *Coronary artery calcium scoring (CACS) as a diagnostic test for detection of coronary artery disease.* Olympia, WA: Health Technology Assessment Program. Retrieved August 31, 2012, from http://www.hta.hca.wa.gov/documents/calcium_scoring_report_final_090409.pdf

The summary of evidence in this document is derived directly from these evidence sources, and portions are extracted verbatim.

Clinical Background

Coronary artery calcification is part of the development of atherosclerosis. It is an active process that begins as early as the second decade of life and occurs exclusively in atherosclerotic arteries and is absent in the normal vessel wall. A close relationship has been confirmed between the extent of coronary artery calcification and the atherosclerotic plaque burden seen in coronary artery disease (CAD), making calcium a potential marker for diseased arteries.

Coronary calcification is pervasive in patients with confirmed CAD and increases with age. Increasing prevalence of coronary artery calcified plaque parallels the increasing prevalence of coronary atherosclerosis over the lifespan. However, the presence of calcified coronary plaque is not strongly correlated with the extent of histopathologic stenosis. The inner lining of both obstructed and non-obstructed vessels contains coronary artery calcified plaque; therefore, the detection of calcified plaque on cardiac CT is not specific to an obstructive lesion.

Currently, the most common method for determining coronary artery calcium (CAC) score use computed tomography (CT), either electron beam CT or multidetector CT for the detection and quantification of the amount of coronary artery calcium. However, calcification in vessels may be present in both obstructive and nonobstructive lesions and thus, coronary artery calcium is not specific for obstructive CAD.

The role of coronary artery calcium scoring (CACS) as a diagnostic or clinical decision-making tool in symptomatic persons has not been well defined. It is not likely to be a replacement for conventional coronary angiography, which is the gold standard anatomical test for CAD. Some proponents of CACS suggest that it may be most useful in separating persons who are unlikely to have significant coronary artery obstruction from those who should be referred for additional diagnostic testing. From this perspective, those with little or no calcium are less likely to have CAD requiring further evaluation, hospitalization or intervention. Those with a positive CACS are then often referred for stress tests to evaluate myocardial function, perfusion studies and/or invasive conventional coronary angiography and appropriate treatment.

In clinical practice, CACS may be used to determine whether patients presenting with chest pain should have further testing. Coronary artery calcium scoring as a stand-alone diagnostic test, however, is less common. (The more common use appears to be the evaluation of asymptomatic patients.) Coronary artery calcium scoring is increasingly performed in conjunction with CT coronary angiography using multidetector CT.

Electron beam CT and multidetector CT, both used for CACS, expose the patient to ionizing radiation. Potential adverse health effects associated with radiation exposure may be of concern to patients as well as clinicians. Presumably patients with a positive CACS may also have other diagnostic tests that involve ionizing radiation. Thus, radiation exposure related to CACS should be put in the context of additional testing that may be indicated.

Evidence Review

US Preventive Services Task Force Report on Using Nontraditional Risk Factors in Coronary Heart Disease Risk Assessment

The report pertains only to asymptomatic patients, and makes the following recommendations:

Clinicians should use the Framingham model to assess coronary heart disease (CHD) risk and to guide risk-based therapy until further evidence is obtained. Because adding nontraditional risk factors (including CACS) to CHD assessment requires additional patient and clinical staff time and effort, routinely screening with nontraditional risk factors could result in lost opportunities for provision of other important health services of proven benefit.

This recommendation is to be used for those who fall into a 10% to 20% (intermediate) 10-year risk category after being screened for CHD risk by using traditional CHD risk factors. Using a risk assessment tool is a key step in managing CHD risk in patients. One validated method of assessing CHD risk is the Framingham model. Persons with low (<10%) Framingham risk scores do not benefit from aggressive risk factor modification, whereas those with high (>20%) Framingham risk scores do benefit. Examples of persons who fall into the intermediate-risk category include a 60-year-old male smoker with untreated hypertension or a 60-year-old female with untreated hypertension and hyperlipidemia. The current recommendation used the Adult Treatment Panel III Framingham risk calculator (available at http://hp2010.nhlbihin.net/atpiii/calculator.asp?usertype.prof) and does not include diabetic populations.

The USPSTF found no evidence that risk stratification with any nontraditional risk factors including CACS, either independently or in addition to Framingham risk scoring, reduces myocardial infarction or cardiovascular disease mortality compared with risk stratification and treatment on the basis of Framingham scoring alone. Therefore, the USPSTF examined the evidence for the independent and additive predictive value of each nontraditional risk factor in assessing 10-year risk for myocardial infarction and CHD mortality. For those risk factors for which evidence for independent or additive

DRAFT

predictive value is available, the USPSTF evaluated the evidence for the effect such factors may have on recategorizing intermediate-risk persons into low- or high-risk groups.

Regarding CACS, the evidence review found poor- to fair-quality evidence indicating that higher CAC scores on electron beam CT predict CHD events independent of Framingham risk factors, on the basis of a systematic review of eight cohort studies. Three good-quality population cohort studies and five fair-quality studies reported that the highest CAC score groups had significantly greater relative risk estimates than the lowest score groups. Although three of the studies met the technical requirements for a good-quality rating, none of them make a convincing case that CAC adds information about intermediate-risk persons. One of the three included only low-risk persons. Another study, from the Rotterdam Coronary Calcification Study, used self-selected participants who were classified into two categories (10-year Framingham risk of >20% or <20%), and results for the intermediate-risk group (10% to 20%) were therefore not reported separately. Several features of the third study, from the South Bay Heart Watch, limit its applicability to an intermediate risk group. The predictive value of a high CAC score was inconsistent; for example, participants with a Framingham risk score of 11% to 15% and participants with a risk score of 16% to 20% had the same baseline risk (7%). The CAC score also seemed to be imprecise; among participants who had a high CAC score, those with a pretest Framingham risk score of 10% to 15% had a higher posttest risk (19%) than those with a pretest score of 16% to 20%. Finally, participants were potentially self-selected. The five studies rated as fair quality were primarily limited by their use of proxy measures to control for Framingham risk factors or their recruitment of self-selected participants.

In summary, although the eight included studies consistently reported statistically significant relative risks for coronary events with increasing CAC scores, no study uniformly met all three of the following conditions: addressed an intermediate-risk cohort, was population-based or free of selection bias, and appropriately measured or controlled for traditional risk factors.

Hayes Report on Use of CACS in Asymptomatic Adults

The available evidence suggests that CACS adds incremental predictive value over traditional risk factor assessments such as the Framingham Risk Score, particularly among asymptomatic adults at intermediate risk of a CAD event. Among three studies, 20% to 55% of those initially classified as intermediate risk were reclassified once CAC scores were considered. However, it is not yet known whether the addition of CACS to standard risk factor assessment will improve patient-important outcomes (i.e., cardiac events). The one randomized trial comparing scanning with conventional risk factor analysis alone reported that CAC scanning was associated with some improvement in clinical risk factors for CAD, but there was no difference in adverse event rate between the scanned and non-scanned groups. Computed tomography-induced radiation exposure is the single biggest safety concern in relation to CACS.

DRAFT

Washington HTA Report (Coronary Artery Calcium Scoring)

The Washington HTA report addresses the use of CACS in symptomatic patients only.

CACS test characteristics
The role of CACS as a diagnostic test is not clear from the literature and there is no consensus on appropriate thresholds for determining a negative versus positive test. It is not likely to be a replacement for conventional coronary angiography based on test performance characteristics. Some literature suggests that it might be used for triaging symptomatic patients (both stable outpatients, and patients with acute chest pain presenting to the emergency department) and that CACS may reduce the use of conventional coronary angiography.

- A CACS > 0 is highly sensitive (99%, CI = 98% - 99%) for identifying the presence of obstructive CAD, however specificity was only 35%.
- At thresholds of CAC scores ≥ 100 (5 studies) or ≥ 400 (3 studies) the sensitivity is lower (85% and 78% respectively) but specificity is improved (77% and 83%, respectively).

Safety of CACS
The primary safety concerns for CACS relate to radiation exposure and the consequences of incidental findings.

- Radiation exposure
  - To date, no large-scale epidemiologic studies evaluating cancer risk associated with CT in general have been published.
  - There is uncertainty and controversy with regard to the actual risk of low dose radiation. Quantification of risk specific to CACS for an individual patient is not possible.
  - A typical effective dose for CACS is estimated to be 3 mSV (reported range 0.7 -12 mSv) when retrospective and prospective gating are considered together. Exposure is less when scans are prospectively gated. Some experts consider the potential for harm from radiation exposure to be clinically significant particularly given that patients may be likely to have additional tests using radiation.
  - A recent simulation estimating radiation dose and cancer risk suggests that a single scan for CACS may increase lifetime cancer risk. For a single screen at 55 years of age, based on a median effective dose of 2.3 mSv, site-specific estimates for lifetime risk of radiation induced cancer suggest that most cases would be lung cancer (6/100,000 in men, 14/100,000 in women) or breast cancer (4/100,000 in women).
  - The extent to which CACS is an adjunct to coronary CT angiography may increase radiation exposure compared with that for CACS alone.
- Consequences of incidental findings

DRAFT

- Data from two studies suggests that 7% to 10% of symptomatic persons will have incidental findings during a CT scan for calcium scoring that require further diagnostic testing and a small percent, 1.2%, will require therapeutic intervention. There may be benefits to early detection and treatment of the small percentage of significant pathology found incidentally, however, there is no evidence from these studies that early detection prompted more effective treatment or enhanced patient outcomes.
- The follow-up of less serious findings may create patient anxiety in addition to exposing them to the inconvenience, costs and risks of additional testing.

Influence on clinical decision making and patient outcomes

- There is an association between CACS and future events: Patients with higher CACS may experience more cardiac events (e.g. myocardial infarction, revascularization, death) and those with no calcium or low scores may be less likely to have future events. The extent to which CACS truly influences outcomes is unclear, however, since its impact on clinical decision making and treatment is not described.
- While there are a number of studies describing the potential role of CACS as a triage tool for ruling out CAD and identifying those who should have additional testing, none of the studies included a comparison group. If CACS was a perfectly sensitive test, there were no false negatives and some degree of specificity, the benefit of doing CACS as a first test for triage could be estimated in the absence of an explicit comparison group. Without this or a comparison group, it is difficult to assess the incremental benefit of CACS in clinical decision making.

Special populations

- Two moderate quality validation studies in symptomatic diabetic patients suggest that the sensitivity (98-99%) and specificity (25-39%) of CACS for the detection of any calcium is similar to that for general populations from the meta-analysis of Level of Evidence (LoE) I/II studies but that a higher percent (11-25%) of persons with a negative test would have CAD.
- Three moderate quality (LoE II/III) studies described performance characteristics for men and women separately. At a CACS >0, the sensitivities for both groups were 96%-100%. Specificities for women ranged for 41% to 66% and those for men 24% to 57%, somewhat lower. A higher percent (4-11%) of men with a negative test would have CAD compared with women (0-4%). The prevalence of CAD was lower in women (36-47%) compared with men (53-70%). Women present with CAD at an older age (~10 years) than men, which may account for the differences.
- Seven LoE I/II studies explored the relationship of age with test performance characteristics. The prevalence of CAD and presence of calcium increases with age. There are, however somewhat mixed results regarding the extent to which age influences test performance characteristics. While some studies suggest that

DRAFT

sensitivity and predictive values go up with increasing age, others suggest that the best sensitivity and specificity may be in middle aged patients (40-60 years).

Economic implications

- Two full economic studies and one costing evaluate CACS as a stand-alone test compared with conventional angiography.
- The two moderate quality full economic studies suggest that at a disease prevalence of up to 70%, CACS may be more cost effective than conventional angiography, however incremental cost effectiveness is not described.
- Disease prevalence and CAC score cut-off (and corresponding sensitivity and specificity) appear to influence overall cost-effectiveness.
- Models did not include evaluation of incidental findings and the influence of false-negative and false-positive tests is not clear.
- Coronary artery calcium scoring does not appear to function as a stand-alone test in clinical practice. The potential impact of additional testing done in clinical practice needs to be considered and modeled.
- There is insufficient evidence for conclusions on the long-term cost utility of CACS compared with conventional coronary angiography alone or with regard to other non-invasive tests.

WA HTA Clinical Committee Decision

The WA HTA clinical committee decided against coverage of CACS. Their rationale is outlined below:

- The committee agreed with the evidence report and found that CACS sensitivity and reliability are high for CACS, though specificity is low and like other tests, accuracy is affected by the disease prevalence. While accuracy and reliability are critical, they are only a first step as to whether a test is effective. The committee also agreed that there is no evidence to establish a clinically important threshold: increase in calcium does indicate disease, but the correlation to severity of stenosis is not established – which is key in a disease that is widely prevalent, where serious events occur in some, but are difficult to predict.
- In evaluating effectiveness, the most rigorous question is whether substituting this test, instead of a current diagnostic, results in better treatment and outcomes. In this case, the evidence is insufficient and current clinical practice does not support using this test alone or as a substitute.
- The other diagnostic effectiveness key question discussed by the committee is whether there is evidence that using this test as an added tool to current strategy provides a benefit (clinical or cost). The remaining analysis relate to answering this question.
- One potential use would be in ER where symptomatic patient at low to intermediate risk - could rule out disease. This use would require CACS of 0 value, so the specificity goes down, and at least a 5% group would still receive a

DRAFT

negative test, but would have disease. One small retrospective study looked at 4 month follow up on 100 patients in ED where CACS score was taken, along with other tests and concluded that a score of 0 could permit a discharge. CACS studies did not include any RCT or higher quality observational trials to explicitly test what different clinical or treatment choices are made. The clinical expert noted that there is usually a need for a functional test to confirm.

- The committee noted that national guidelines do not endorse the use of CACS, though some have permissive statements for use of the test.

NICE Guideline: Chest Pain of Recent Onset

The NICE guideline does not address the use of CACS in patients presenting with acute chest pain. For patients presenting with stable chest pain in the outpatient setting, they make the following recommendations pertaining to CACS:

> In people without confirmed CAD, in whom stable angina cannot be diagnosed or excluded based on clinical assessment alone, estimate the likelihood of CAD (see Table 1). Take the clinical assessment and the resting 12-lead ECG [electrocardiogram] into account when making the estimate. Arrange further diagnostic testing as follows:
>
> - If the estimated likelihood of CAD is 61–90%, offer invasive coronary angiography as the first-line diagnostic investigation if appropriate.
> - If the estimated likelihood of CAD is 30–60%, offer functional imaging as the first-line diagnostic investigation.
> - If the estimated likelihood of CAD is 10–29%, offer CACS as the first-line diagnostic investigation. If the calcium score is:
> - zero, consider other causes of chest pain
> - 1–400, offer 64-slice (or above) CCTA
> - greater than 400, offer invasive coronary angiography.

DRAFT

Table 1. Percentage of people estimated to have coronary artery disease according to typicality of symptoms, age, sex and risk factors

	Non-anginal chest pain				Atypical angina				Typical angina			
	Men		Women		Men		Women		Men		Women	
Age (years)	Lo	Hi	Lo	Hi	Lo	Hi	Lo	Hi	Lo	Hi	Lo	Hi
35	3	35	1	19	8	59	2	39	30	88	10	78
45	9	47	2	22	21	70	5	43	51	92	20	79
55	23	59	4	25	45	79	10	47	80	95	38	82
65	49	69	9	29	71	86	20	51	93	97	56	84

For men older than 70 with atypical or typical symptoms, assume an estimate > 90%.
For women older than 70, assume an estimate of 61–90% EXCEPT women at high risk AND with typical symptoms where a risk of > 90% should be assumed.

Values are percent of people at each mid-decade age with significant coronary artery disease (CAD)[1].
Hi = High risk = diabetes, smoking and hyperlipidaemia (total cholesterol > 6.47 mmol/litre).
Lo = Low risk = none of these three.
The shaded area represents people with symptoms of non-anginal chest pain, who would not be investigated for stable angina routinely.
Note: These results are likely to overestimate CAD in primary care populations. If there are resting ECG ST-T changes or Q waves, the likelihood of CAD is higher in each cell of the table.

Discussion of the evidence for CACS in the NICE guideline is as follows:

Advantages and Disadvantages
The main advantages of calcium scoring are that calcium scanning takes approximately five minutes to perform and interpret, there is minimal radiation exposure (1.5 to 3 mSv) compared with multislice coronary angiography, no contrast material is required, the quantification of plaque (calcium score) enables non invasive temporal tracking of atherosclerosis burden and, although not of direct relevance to the investigation of CAD, it detects significant extra-cardiac findings in 2% to 3% as a coincidental finding. The disadvantages include the following; does not assess whether significant coronary stenoses are present, does not make a functional assessment of myocardial ischaemia, and left ventricular function is not assessed. Although coronary artery calcium is well correlated with total plaque volume or atherosclerotic burden it is not a direct marker of the vulnerable plaque at risk of rupture. However, the greater the calcium score the greater the potential for increased numbers of potentially lipid-rich plaques.

Evidence of Diagnostic Efficacy
No systematic reviews were identified. Ten studies were reviewed in total. With increasing thresholds of Agatston calcium score ranges, (from > 0 to 100, and > 100 in 3 studies, and from > 0 to 100, >100 to 400, and > 400 in 3 studies) the sensitivity decreased and the specificity increased for the detection of significant CAD. No evidence was found for the diagnostic accuracy of coronary calcium scores to diagnose significant CAD in ethnic minority groups in the UK. From economic modelling undertaken for this guideline, there is evidence that for patients with a low pre-test-

[1] Adapted from Pryor DB, Shaw L, McCants CB et al. (1993) Value of the history and physical in identifying patients at increased risk for coronary artery disease. *Annals of Internal Medicine 118*(2),81-90.

probability of CAD (<25%), 64-slice CT coronary angiography preceded by testing using calcium scoring is cost-effective compared to functional testing and invasive coronary angiography.

Economic Evaluations
Of the six economic evaluations included in evidence reviewed for this guideline, only one addressed CACS. Rumberger 1999 compared exercise ECG, stress echocardiography (ECHO), stress thallium and CACS. The incremental analysis showed that electron beam CT using a calcium score threshold of >37, >80 or >168 is cost saving compared with stress ECHO and stress thallium testing. At low to moderate disease prevalence (10% to 20%), electron beam CT using thresholds of >37, >80 or >168 are cost saving compared with exercise ECG. Electron beam CT using a threshold of >0 is cost saving compared with stress thallium testing at 20% CAD prevalence and above.

The NICE guideline authors performed their own economic analysis of a diagnostic strategy that incorporated the use of calcium scoring using 64-slice CT coronary angiography as a precursor to full 64-slice CT coronary angiography. This was done as a way of minimizing the risk of radiation from 64- slice CT coronary angiography, a risk which was not explicitly incorporated into the other models. Results of the base case analysis indicate that for lower risk groups (5% and 20%), the use of calcium scoring as a first line testing strategy is likely to be cost-effective and should be followed by either 64-slice CT coronary angiography alone or with additional invasive coronary angiography as a confirmatory 3rd test. In higher risk populations, (CAD prevalence greater than 40%), a strategy of sending all patients directly to invasive coronary angiography is likely to be cost-effective. The model indicates that MPS with SPECT is excluded through dominance or extended dominance at every level of CAD prevalence. It also indicates that exercise ECG is only cost-effective as a first line investigation strategy at 5% CAD prevalence, but that even in this instance replacing exercise ECG with calcium scoring is likely to improve effectiveness at a reasonable level of additional cost.

Overall Summary

There is no evidence that risk stratification in asymptomatic patients using CACS reduces myocardial infarction or cardiovascular disease mortality compared with risk stratification and treatment on the basis of Framingham scoring alone. Coronary artery calcium scoring may have a diagnostic role in the “rule out” of obstructive CAD in emergency department patients with acute chest pain and normal ECGs and initial cardiac enzymes, and in outpatients with stable chest pain with a low probability of obstructive CAD. However, there is little data available to support long-term outcomes using calcium scoring as a strategy, and it does not appear to function as a stand-alone test in clinical practice. The potential impact of radiation exposure, both from the CACS and from additional testing done to confirm the diagnosis or to evaluate incidental findings, needs to be considered, and current studies do not adequately address these concerns. One economic evaluation suggests that the most cost-effective course of action for stable outpatients with a low probability of CAD (10-29%) is CACS, followed by CCTA if the CACS score is 1-400, or invasive angiography if the score is greater

than 400, however, this was from the perspective of the UK National Health Service, and applicability to the US setting is limited given differences in costs and the non-existence of accepted follow up algorithms.

PROCEDURE

Electron beam coronary computed tomography
Multidetector coronary computed tomography
Coronary Artery Calcium Scoring

DIAGNOSES

Coronary artery disease
Chest pain

APPLICABLE CODES

CODES	DESCRIPTION
ICD-9 Diagnosis Codes	
410	Acute myocardial infarction
411	Other acute and subacute forms of ischemic heart disease
413	Angina pectoris
414	Other forms of chronic ischemic heart disease
ICD-9 Volume 3 (Procedure Codes)	
87.41	Computed axial tomography of the heart
CPT Codes	
75571	Computed tomography of heart, without contrast, with qualitative evaluation of coronary calcium
HCPCS Codes	
None	

Note: Inclusion on this list does not guarantee coverage

Coverage guidance is prepared by the Health Evidence Review Commission (HERC), HERC staff, and subcommittee members. The evidence summary is prepared by the Center for Evidence-based Policy at Oregon Health & Science University (the Center). This document is intended to guide public and private purchasers in Oregon in making informed decisions about health care services.

The Center is not engaged in rendering any clinical, legal, business or other professional advice. The statements in this document do not represent official policy positions of the Center. Researchers involved in preparing this document have no affiliations or financial involvement that conflict with material presented in this document.

Washington State Health Care Authority

Health Technology Assessment

HTA Final Report

Coronary Artery Calcium Scoring (CACS) as a Diagnostic Test for Detection of Coronary Artery Disease

Date: Friday, September 4th, 2009

Health Technology Assessment Program
676 Woodland Square Loop SE
P.O. Box 42712
Olympia, WA 98504-2712
http://www.hta.hca.wa.gov

Coronary Artery Calcium Scoring (CACS) as a Diagnostic Test for Detection of Coronary Artery Disease

Provided by:

Spectrum Research, Inc.

Prepared by:

Andrea C. Skelly, PhD, MPH
Erika D. Ecker, BS
Nora B. Henrikson, PhD, MPH
Carin M. Olson, MD, MS
Annie L. Raich, MS, MPH
Ellen M. Van Alstyne, MS
Joseph R. Dettori, PhD, MPH

Biostatistics assistance from
Margaret S. Pepe, PhD, MS
Daryl Morris, MS

This technology assessment report is based on research conducted by a contracted technology assessment center, with updates as contracted by the Washington State Health Care Authority. This report is an independent assessment of the technology question(s) described based on accepted methodological principles. The findings and conclusions contained herein are those of the investigators and authors who are responsible for the content. These findings and conclusions may not necessarily represent the views of the HCA/Agency and thus, no statement in this report shall be construed as an official position or policy of the HCA/Agency.

The information in this assessment is intended to assist health care decision makers, clinicians, patients and policy makers in making sound evidence-based decisions that may improve the quality and cost-effectiveness of health care services. Information in this report is not a substitute for sound clinical judgment. Those making decisions

regarding the provision of health care services should consider this report in a manner similar to any other medical reference, integrating the information with all other pertinent information to make decisions within the context of individual patient circumstances and resource availability.

TABLE OF CONTENTS

Executive Summary 6

Appraisal 11

1. Background 14
1.1 The Condition 14
1.2 The Technology and its Comparator(s) 16
1.3 Clinical Guidelines 22
1.4 Previous Systematic Reviews/Technology Assessments 25
1.5 Medicare and Representative Private Insurer Coverage Policies 26
1.6 Washington State Data 29

2. The Evidence 32
2.1 Systematic Literature Review 32
2.2 Methods 32
2.3 Quality of Literature Available 38
2.4 Description of Study Population 42
2.5 Description of Study Outcomes 43

3. Results 43
3.1 Key Question 1: What are the test characteristics, PPV (positive predictive value), NPV (negative predictive value), sensitivity and specificity, of coronary artery calcium scoring (CACS) compared with the reference standard of coronary angiography for the diagnosis of CAD or other established diagnostic tests for CAD. What is the evidence to describe the reliability (i.e., test-retest, intra-reader, inter-reader performance) of CACS? 43
3.2 Key Question 2: What is evidence related to the safety of CACS? 56
3.3 Key Question 3: What is the evidence that CACS influences clinical decision making and improves patient clinical outcomes (e.g. mortality)? 62
3.4 Key Question 4: What is the evidence that CACS may perform differently in special populations (e.g. women, diabetic populations)? 67
3.5 Key Question 5: What evidence of cost implications and cost-effectiveness for CACS compared with other diagnostic tests? 72

Summary and Implications 79

References 85

TABLES

Table 1 Overall Strength of Evidence (SoE) Criteria 9
Table 2 Summary of findings and overall strength of evidence 9
Table 3 Summary of payer policies 26
Table 4 Summery of inclusion and exclusion criteria 32
Table 5 Level of evidence (LoE) summary for LoE I/II validation studies 39
Table 6 Population characteristics of LoE I/II validation studies 42
Table 7 Summary of test characteristics at alternate thresholds 51
Table 8 LoE III studies for which data at CACS > 0 were available 51
Table 9 Overview of main findings for reliability studies on coronary artery calcium scoring (CACS) 55
Table 10 Overview of typical effective dose for various radiation sources 57
Table 11 Summary of studies reporting incidental findings on CT for calcium scoring 61
Table 12 Patient characteristics from studies of diabetic patients 68
Table 13 Summary of CACS test characteristics in diabetic populations 68
Table 14 Patient characteristics in male and female study populations 69
Table 15 Summary of CACS performance in male and female populations compared with angiography 70
Table 16 The effect of age on diagnostic validity of calcium scores 71
Table 17 Overview of economic studies for CACS 74
Table 18 Overview of results from full economic studies 76

FIGURES

Figure 1 Algorithm for article selection 33
Figure 2 Flow chart showing results of literature search 34
Figure 3 Meta-analysis of test characteristics for LoE I/II studies reporting data for a CACS threshold of > 0 as a positive test 46
Figure 4 Meta-analysis of test characteristics for LoE I/II studies reporting data for a CACS threshold of ≥ 100 as a positive test 48
Figure 5 Meta-analysis of test characteristics for LoE I/II studies reporting data for a CACS threshold of ≥ 400 as a positive test 50

EXECUTIVE SUMMARY

Introduction

Coronary artery disease (CAD), also referred to as coronary heart disease (CHD), is the single leading cause of death for both men and women in the United States and is the most common form of cardiovascular disease. Particularly in patients with a low pre-test probability of CAD (e.g. younger persons, women), symptoms of CAD have poor specificity and sensitivity for CAD, so diagnostic testing is used to confirm or refute a clinical suspicion of CAD. Coronary artery calcium scoring (CACS) is one such test that may provide diagnostic information to help either rule out or rule in significant CAD and assist with clinical decision making. CACS provides anatomical information on the amount of calcium in the heart and coronary arteries.

Methods for evaluating comparative effectiveness

Spectrum Research, Inc.'s (SRI) method for technology assessment involves formal, structured systematic search of the peer-reviewed literature across a number of databases (e.g. MEDLINE, EMBASE) in addition to searches of pertinent databases related to clinical guidelines and previously performed assessments. Included studies are appraised using SRI's Level of Evidence (LoE) system which evaluates the methodological quality based on study design and factors that may bias studies. An overall Strength of Evidence (SoE) combines the LoE with consideration of the number of studies and consistency of the findings to describe an overall confidence regarding the stability of estimates as further research is available. Included economic studies were also formally appraised based on criteria for quality of economic studies and pertinent epidemiological precepts.

Throughout the process, SRI sought expert input to ensure that the methodological and clinical components are accurately represented and relevant. In addition, peer-review by clinical experts, researchers and those with expertise in economic and outcomes evaluation were invited to provide an assessment of the systematic review methodology, analyses and report conclusions.

Summary and Implications

1. CACS test characteristics: Validation and accuracy, reliability and reproducibility of CACS compared with CCA.

- The role of coronary artery calcium scoring (CACS) as a diagnostic test is not clear from the literature and there is no consensus on appropriate thresholds for determining a negative versus positive test. It is not likely to be a replacement for conventional coronary angiography (CCA) based on test performance characteristics. Some literature suggests that it might be used for triaging symptomatic patients and that CACS may reduce the use of conventional coronary angiography.
- Based on meta-analysis of LoE I/II studies comparing CACS with the reference standard of conventional coronary angiography, the overall strength of evidence is high.
    - A CACS > 0 is highly sensitive (99%, CI = 98% - 99%) for identifying the presence of obstructive CAD, however specificity was only 35% and 5% of persons (1 – negative predictive value) with a negative test would have

CAD based on pooled estimates from seven studies with a total of N = 7354 patients. Approximately 35% of persons without CAD might avoid unnecessary angiography or additional tests.

- At thresholds of ≥ 100 (5 studies) or ≥ 400 (3 studies) the sensitivity is lower (85% and 78% respectively) but specificity is improved (77% and 83%, respectively). Clear decisions may not be possible based on CACS when using these thresholds to define a positive test.

2. Safety of CACS

The primary safety concerns for CACS relate to radiation exposure and the consequences of incidental findings.

Radiation exposure

- The overall strength of evidence regarding safety is very low primarily due to uncertainties regard the cancer-related risks due to radiation exposure particularly when CACS may lead to additional tests involving radiation. On the other hand, to the extent that CACS has the potential to decrease the use of conventional angiography, overall radiation exposure might be reduced.
- To date, no large-scale epidemiologic studies evaluating cancer risk associated with computed tomography (CT) in general have been published.
- There is uncertainty and controversy with regard to the actual risk of low dose radiation. Quantification of risk specific to CACS for an individual patient is not possible.
- A typical effective dose for CACS is estimated to be 3 mSV (reported range 0.7 - 12 mSv) when retrospective and prospective gating are considered together. Exposure is less when scans are prospectively gated. Some experts consider the potential for harm from radiation exposure to be clinically significant particularly given that patients may be likely to have additional tests using radiation.
- A recent simulation estimating radiation dose and cancer risk suggests that a single scan for CACS may increase lifetime cancer risk. For a single screen at 55 years of age, based on a median effective dose of 2.3 mSv, site-specific estimates for lifetime risk of radiation induced cancer suggest that most cases would be lung cancer (6/100,000 in men, 14/100,000 in women) or breast cancer (4/100,000 in women).
- Decision making between physician and patient should involve a discussion of the potential risks and benefits of CACS (and subsequent testing). Final determination of net benefit for a given clinical scenario reflects the values and judgments of the persons making the decisions.
- The extent to which CACS is an adjunct to coronary CT angiography may increase radiation exposure compared with that for CACS alone.

Consequences of Incidental findings

- The overall strength of evidence is very low.
- Data from two studies suggests that 7%-10% of symptomatic persons will have incidental findings during a CT scan for calcium scoring that require further diagnostic testing and a small percent, 1.2%, will require therapeutic intervention.

- There may be benefits to early detection and treatment of the small percentage of significant pathology found incidentally, however, there is no evidence from these studies that early detection prompted more effective treatment or enhanced patient outcomes.
- The follow-up of less serious findings may create patient anxiety in addition to exposing them to the inconvenience, costs and risks of additional testing.

3. Influence on clinical decision making and patient outcomes

- There is an association between CACS and future events: Patients with higher CACS may experience more cardiac events (e.g. myocardial infarction, revascularization, death) and those with no calcium or low scores may be less likely to have future events. The extent to which CACS truly influences outcomes is unclear, however, since its impact on clinical decision making and treatment is not described.
- Overall, the evidence is low that CACS facilitates clinical decision making. While there are a number of studies describing the potential role of CACS as a triage tool for ruling out CAD and identifying those who should have additional testing, none of the studies included a comparison group. If CACS was a perfectly sensitive test, there were no false negatives and some degree of specificity, the benefit of doing CACS as a first test for triage could be estimated in the absence of an explicit comparison group. Without this or a comparison group, it is difficult to assess the incremental benefit of CACS in clinical decision making.

4. Special populations

- Two moderate quality validation studies in symptomatic diabetic patients suggest that the sensitivity (98-99%) and specificity (25%-39%) of CACS for the detection of any calcium is similar to that for general populations from the meta-analysis of LoE I/II studies but that a higher percent (11%-25%) of persons (1 – negative predictive value) with a negative test would have CAD. The overall strength of evidence is very low.
- Three moderate quality (LoE II/III) studies described performance characteristics for men and women separately. At a CACS >0, the sensitivities for both groups were 96%-100%. Specificities for women ranged for 41%-66% and those for men 24%-57%, some what lower. A higher percent (4% - 11%) of men (1 – negative predictive value) with a negative test would have CAD compared with women (0%-4%). The prevalence of CAD was lower in women (36%-47%) compared with men (53%-70%) Women present with CAD at an older age (~10 years) than men, which may account for the differences.
- Seven LoE I/II studies explored the relationship of age with test performance characteristics. The prevalence of CAD and presence of calcium increases with age. There are, however somewhat mixed results regarding the extent to which age influences test performance characteristics. While some studies suggest that sensitivity and predictive values go up with increasing age, others suggest that the best sensitivity and specificity may be in middle aged patients (40 – 60 years). The overall strength of evidence for studies with regard to age is moderate.

5. Economic implications

- Two full economic studies and one costing evaluate CACS as a stand-alone test compared with conventional angiography.
- The two moderate quality full economic studies suggest that at a disease prevalence of up to 70%, CACS may be more cost effective than conventional angiography, however incremental cost effectiveness is not described.
- Disease prevalence and CACS score cut-off (and corresponding sensitivity and specificity) appear to influence overall cost-effectiveness.
- Models did not include evaluation of incidental findings and the influence of false-negative and false-positive tests is not clear.
- CACS does not appear to function as a stand-alone test in clinical practice. The potential impact of additional testing done in clinical practice needs to be considered and modeled.
- There is insufficient evidence for conclusions on the long-term cost utility of CACS compared with CCA alone or with regard to other non-invasive tests.

Table 1. Overall Strength of Evidence (SoE) Criteria

SoE	Description	Further Research Impact	Domain Criterion Met		
			Quality	Quantity	Consistency
1	High	Very unlikely to change confidence in effect estimate	+	+	+
2	Moderate	Likely to have an important impact on confidence in estimate and *may* change the estimate	+	-	+
			+	+	-
3	Low	Very likely to have an important impact on confidence in estimate and *likely* to change the estimate	+	-	-
			-	+	+
4	Very Low	Any effect estimate is uncertain	-	+	-
			-	-	+
			-	-	-

Table 2. Summary of findings and overall strength of evidence

Key Question 1: Evidence regarding test characteristics and reliability		
Outcome	**Strength of Evidence**	**Results**
Validity of test	1	The role of CACS as a stand alone diagnostic test is not clear. There is no consensus on threshold. Based on meta analysis of LoE I/II studies • A CACS > 0 is highly sensitive (99% , CI = 98% - 99%) for identifying the presence of obstructive CAD, however 5% of persons (1 – negative predictive value) with a negative test would have CAD • At thresholds of ≥ 100 (5 studies) or ≥ 400 (3 studies) the sensitivity is lower (85% and 78% respectively) but specificity is improved (77% and 83%, respectively)
Reliability of test	1	• The reliability of CACS (based on Agaston method) appears to be moderate to high based on 3 small LoE II studies and descriptions in it two validation studies
Key Question 2: Evidence regarding safety		
Radiation	4	• While simulation and modeling of the effects of radiation exposure provide important insights into the possible changes in risks, the true attributable risk from radiation-based diagnostic tests may be difficult to determine. • Radiation exposure may be reduced to the extent that CACS use avoids doing angiography. On the other hand, exposures may be increased to the extent that positive

		CACS results in additional testing. • A typical effective dose for CACS is estimated to be 3mSV (reported range 1-12mSv). CACS results may lead to additional testing which involves radiation. • In a recently published simulation based on a median effective dose of 2.3 mSv, site-specific estimates for life-time risk of radiation-induced cancer suggest that most cases would be lung cancer (6/100,000 in men, 14/100,000 in women) or breast cancer (4/100,000 in women). • Decision making should include discussion of the potential for such risks.
Incidental findings	**4**	• 7%-10% of symptomatic persons will have incidental findings during a CT scan for calcium scoring that require further diagnostic testing and a small percent, 1.2%, will require therapeutic intervention based on two studies in symptomatic persons.
Key Question 3: Evidence regarding clinical decision making and patient outcomes		
Triage in emergency department	**3**	• Five studies suggest that a CACS = 0 may allow discharge of patients with suspected CAD. These studies, however vary in quality. None employed a comparison group and are considered case series.
Triage in other clinical settings	**4**	• One study reported that referral to conventional angiography increased with increasing CACS. No comparison group was employed.
Prediction of future events	**3**	• While 3 studies suggest that CACS is a predictor of future cardiac events, none evaluate the role of therapeutic interventions which may influence the occurrence of such events.
Key Question 4: Evidence regarding performance in special populations		
Diabetes	**4**	• Sensitivity (98-99%) and specificity (25%-39%) of CACS for the detection of any calcium is similar to that for general populations from the meta-analysis of LoE I/II studies but a higher percent (11%-25%) of persons (1 – negative predictive value) with a negative test would have CAD based on two moderate quality studies.
Gender	**3**	• Three studies evaluated CACS characteristics in women vs. men. Sensitivities were similar for both groups at CACS > 0. Specificities for women ranged for 41%-66% and those for men 24%-57%, some what lower. • A higher percent (4% - 11%) of men (1 – negative predictive value) with a negative test would have CAD compared with women (0%-4%)., however, the prevalence of CAD was lower in women (36%-47%) compared with men (53%-70%) • Women present with CAD at an older age (~10 years) than men, which may account for the differences
Age	**2**	• Seven LoE I/II validation studies evaluated the influence of age on CACS. In general, the prevalence of coronary artery calcium increases with age. • There are conflicting results regarding test performance at various ages.
Key Question 5: Evidence regarding cost-effectiveness		
	4	• Two moderate quality studies suggest that at a disease prevalence of up to 70%, CACS may be more cost effective than conventional angiography, however incremental cost effectiveness is not described. • Cost-effectiveness is influenced by disease prevalence and CACS score cut-off (and corresponding sensitivity and specificity) • The influence of additional testing to reflect clinical practice needs to be more fully considered. • The influence of false-negative and false positive results is unclear and models did not consider follow-up of incidental findings. • There is insufficient evidence for conclusions on the long-term cost utility of CACS compared with CCA alone or with regard to other non-invasive tests.

APPRAISAL

Coronary artery calcium scoring (CACS)

Final Scope

Rationale for the Appraisal

The role of coronary artery calcium scoring (CACS) as a diagnostic test and component of clinical decision making in symptomatic persons in whom coronary artery disease (CAD) is suspected is not well established. Questions remain about the role of CACS as a diagnostic test. Thresholds for the amount and types of plaque that may signify CAD that requires intervention are not well delineated. It is also not clear whether coronary artery calcium (CAC) detection and scoring changes treatment decisions. Evidence regarding which persons might benefit from diagnostic testing and whether early intervention provides better health outcomes or leads to additional unnecessary interventions, especially invasive interventions, is not well described. From a public health perspective, a diagnostic test should only be performed if it leads to the use of interventions that, on average, are likely to improve patient outcomes or if it prevents the use of interventions that are not likely to improve outcomes.

Currently, computed tomography (CT) techniques (electron beam or multi-detector) are the most common methods for determining CACS. In symptomatic persons, CACS has been studied as a noninvasive, indirect method for determining obstructive CAD. The presence of calcium is not specific for obstructive CAD since it may be present in both obstructive and non-obstructive lesions. Conventional coronary angiography (CCA) is considered the definitive method for determining presence of obstructive CAD, despite its limitations and potential for subjectivity. It is an invasive test which has associated risks and higher costs. Proponents of CACS suggest that it may facilitate clinical decision-making by identifying persons with low probability of significant coronary artery stenosis, who may not need further evaluation, and for identifying those who should go on for further evaluation.

Objective

The primary aim of this assessment is to systematically review, critically appraise and summarize research evidence describing characteristics of CACS as a *diagnostic* test for evaluation of CAD in patients in whom CAD is suspected. Available information on the economic impact of this will also be summarized and critically appraised. The use of CACS as a screening test in asymptomatic persons is not addressed in this report.

Key questions

When used to diagnose persons with suspected coronary artery disease (CAD):

1. What are the test characteristics, PPV (positive predictive value), NPV (negative predictive value), sensitivity and specificity, of coronary artery calcium scoring (CACS) compared with the reference standard of coronary angiography for the diagnosis of CAD or other established diagnostic tests for CAD. What is the evidence to describe the reliability (i.e., test-retest, intra-reader, inter-reader performance) of CACS.

2. What is the evidence regarding the safety of CACS?
3. What is the evidence that CACS influences clinical decision making and improves patient clinical outcomes (e.g. mortality)?
4. What is the evidence that CACS may perform differently in special populations (e.g. women, diabetic populations)?
5. What evidence of cost implications and cost-effectiveness for CACS compared with other diagnostic tests?

Outcomes

The primary focus of this report is on diagnostic test performance characteristics that describe the validity and accuracy CACS as a diagnostic tool in symptomatic persons when compared with conventional coronary angiography (CCA; the most appropriate reference standard). Test characteristics that have been included are: sensitivity, specificity, positive predictive value and negative predictive value. Information on the reliability (test- retest, intra- and inter-rater reliability) of CACS and information on cost per correct diagnosis from published economic studies are summarized as well.

Key considerations highlighted by clinical experts:

1. **Interventions**

Currently, the most common methods for determining CACS use computed tomography (CT), either electron beam CT (EBCT) or multidetector CT (MDCT) for the detection and quantification of the amount of coronary artery calcium. While there is a correlation between the site and amount of coronary artery calcium with vessel narrowing at that site, as with angiographic stenosis, the vulnerability of the plaque and probability of rupture is not known. In addition, calcification in vessels may be present in both obstructive and nonobstructive lesions and thus, CAC is not specific for obstructive CAD.[1]

The role of CACS as a diagnostic or clinical decision-making tool in symptomatic persons has not been well defined. It is not likely to be a replacement for CCA, which is the gold standard anatomical test for CAD. Some proponents of CACS suggest that it may be most useful in separating persons who are unlikely to have significant coronary artery obstruction from those who should be referred for additional diagnostic testing. From this perspective, those with little or no calcium are less likely to have CAD requiring further evaluation, hospitalization or intervention. Those with a positive CACS are then often referred for stress tests to evaluate myocardial function, perfusion studies and/or invasive CCA and appropriate treatment. The current ACC/AHA guidelines indicate that non-invasive testing for ischemia be done prior to consideration for revascularization. Since CACS is anatomical test and doesn't detect ischemia, it is likely not a replacement for other non-invasive testing which evaluates cardiac and myocardial function.

In clinical practice, CACS may be used to determine whether patients presenting with chest pain should have further testing. CACS as a stand-alone diagnostic test, however, is less common. (The more common use appears to the evaluation of asymptomatic

patients.) CACS is increasingly performed in conjunction with CT coronary angiography using MDCT.

EBCT and MDCT, both used for coronary artery calcium scoring, expose the patient to ionizing radiation. Potential adverse health effects associated with radiation exposure may be of concern to patients as well as clinicians. Presumably patients with a positive CACS may also have other diagnostic tests that involve ionizing radiation. Thus, radiation exposure related to CACS should be put in the context of additional testing that may be indicated. If CACS is done in combination with CT angiography and/or with newer MDCT, the exposure is likely to be greater than with CACS alone using EBCT. The potential risks and benefits with regard to such procedures should be discussed with patients. If use of CACS decreases the need for conventional coronary angiography, radiation exposure may be less.

2. **Costs**

The direct procedural costs of CACS are less than that of CCA. Indirect costs related to patient time are also lower.

The precise role of CACS in symptomatic patients is unclear and clinicians may not be comfortable making decisions on the basis of CACS alone. Consequently, additional tests are often done in clinical practice. A positive CACS frequently leads to stress testing or coronary angiography. These factors influence the evaluation of cost-effectiveness as well as radiation exposure and overall risk/benefit.

3. **Patient considerations**

The pre-test likelihood of CAD is largely based on sex, age, type of chest pain as well as other symptoms and factors. This is important when considering use of a diagnostic test and in comparing the results of studies validating tests such as CACS. Predictive values are influenced by the pre-test probability of disease.

Proponents of CACS suggest that a score of 0 (no coronary artery calcified plaque) in symptomatic patients has high negative predictive values, indicating it may be useful in excluding obstructive angiographic CAD.[2] The predictive values, however, are influenced by a patient's pretest likelihood of disease. In patients with low pre-test probability of CAD, a CACS of zero is associated with low risk of CAD. This may not be the case, however, in patients with higher pre-test likelihoods based on risk factors and clinical symptoms i.e. a negative test may not "rule out" disease to the same extent that it does when the pre-test likelihood is low. [3, 4] In a large study assessing the interrelation of CACS and inducible ischemia in patients with intermediate likelihood of CAD, Schenker, et al reported that 16% of patients with CACS = 0 had myocardial ischemia on provocative testing.[4]

Conventional angiography is an invasive test that requires sedation and carries with it certain risks. CACS is a noninvasive test that can be completed in around 15 minutes and requires no sedation, hospital stay or recovery; it requires only a few seconds of actual scanning time.

A recent simulation study modeling lifetime risk of radiation-induced cancer suggests that even a single scan for CACS determination may increase this risk. Although the typical effective radiation dose is less than that with conventional coronary angiography, the dose is highly variable depending on the equipment and protocol. A 10-fold variation in effective dose has been reported in the literature. There are no standardized protocols for quantification of CACS using MDCT.[5] Anecdotally, the majority of CACS is currently done with prospective gating which decreases exposure. The ACCF/AHA documents provide guidelines for reducing radiation exposure.

Since scanning includes not only the coronary arteries but surrounding structures as well, potential abnormalities in structures other than the coronary arteries and heart might be observed and further evaluation may be required.

4. **Professional considerations:**

CACS currently is most frequently done with a multidetector or electron beam CT scanner, which are expensive pieces of equipment. Sales of multidetector CT scanners are rising with the increasing popularity of cardiac CTA. The CACS requires minimal physician time as the tests are performed by a technician, read by a computer algorithm, and then confirmed by a physician.

5. **Ethical considerations**

Although CACS is a noninvasive test, it does expose the patient to ionizing radiation, a factor which clinicians need to consider and put in the context of other tests that may be part of the clinical pathway which also may use ionizing radiation. CACS often leads to additional testing. Increased use of MDCT over EBCT and combining of CACS with CT angiography may increase patient exposure to ionizing radiation. The benefits, risks and costs therefore need to be considered and discussed with the patient.

1. Background

1.1 The Condition - Coronary Artery Disease

Coronary artery disease (CAD), also referred to as coronary heart disease (CHD), is the number one cause of death and disability in the United States for both men and women, affecting more than 16 million Americans, and is the most common form of cardiovascular disease. Each year, CAD kills more Americans than cancer.[6] In 2005, 652,091 people died of cardiovascular disorders, accounting for 27.1% of all U.S. deaths, with 445,687 (68.3%) of those deaths due to CAD alone.[6, 7] Globally, CAD killed more that 7.6 million people in 2005, which is roughly 43.2% of all cardiovascular-related deaths worldwide.[8] The increasing prevalence and burden of heart disease is also reflected economically. The total of direct and indirect costs of CAD in 2006 was $142.5 billion, with $11.6 billion paid to Medicare beneficiaries ($11,308 per hospital discharge for coronary atherosclerosis).[9] Reduction in the prevalence, morbidity and mortality

related to CAD is an important public health goal given the significant disease burden and contribution to total health care costs.

The underlying cause of CAD is atherosclerosis, a systematic disease process in which plaque, comprised of fat, cholesterol, calcium, and other substances found in the blood, builds up within the walls of damaged arteries leading to hardening or narrowing of the vessels. The coronary arteries, which supply blood, oxygen, and vital nutrients to the heart, can become partially or completely blocked due the build up of plaque. The blocking of the coronary arteries leads to restricted blood flow (ischemia) to the myocardium, weakening it or even causing cell death. Coronary artery plaques are responsible for over 90% of ischemic heart disease (IHD).[10] Common symptoms that occur with CAD are chest pain (angina), arrhythmias, shortness of breath (dyspnea), and in the event of a complete blockage, heart attack. Acute coronary syndromes, such as myocardial infarction (MI) and unstable angina, arise from rupture or erosion of atherosclerotic plaques. Common risk factors for CAD include smoking, high cholesterol, high blood pressure, insulin resistance or diabetes, obesity, metabolic syndrome, sedentary lifestyle, age, and genetics.

CAD develops slowly over time and may be asymptomatic for many years. The onset of symptoms depends on the location and severity of these obstructions; however, the severity of the lesions is poorly correlated with symptoms. Thus, the appropriate diagnostic evaluation of patients who may have CAD is of clinical and practical importance.

Assessment of coronary artery disease

The pre-test likelihood of CAD is largely based on sex, age, type of chest pain as well as other symptoms and factors. The first step in CAD diagnosis is a history and physical which includes consideration of these.

Sensitivity and specificity are the most widely used statistics to describe the accuracy of a diagnostic test. A sensitive test correctly identifies disease in those people who are truly sick, and a test that is specific for a particular disease correctly identifies those people who are well as not having the disease. Ideally, a diagnostic test should achieve both a very high sensitivity and a very high specificity; in reality, however, there is usually a trade off between sensitivity and specificity such that only one can be maximized at a time.

Based on the patient's history and physical, the first diagnostic test to evaluate coronary artery disease (CAD) in patients presenting with signs or symptom may be a simple and noninvasive test called an exercise treadmill test (ETT) which allows for the evaluation of ECG changes in response to exercise. The test is relatively inexpensive, readily available, and the test does not require exposure to radiation. However, wide variability in the diagnostic accuracy of treadmill stress testing as been reported.[11] If the results are ambiguous, imaging studies such as a nuclear SPECT (single photon emission computed tomography) perfusion study or a stress echocardiogram may be ordered. Both of these tests have increased accuracy for diagnosis of CAD compared with ETT. They have the

added advantage of allowing localization of ischemia and do not require the patient to exercise, which is beneficial for those patients who are not ambulatory.[1]

When further evaluation is warranted, or in the acute setting, coronary angiography, a minimally invasive procedure that uses x-ray imaging to visualize coronary anatomy, may be performed. Coronary angiography is performed by inserting a long, thin, flexible tube called a catheter into the body through the groin, arm, or neck, which is then threaded into the coronary arteries where a dye is injected into the bloodstream. While the dye is flowing through the coronary arteries, an x-ray machine rapidly takes a series of images, offering a detailed look at blood flow through the coronary arteries. This procedure can help recognize and treat various disorders including occlusion due to the build up of plaque, stenosis, thrombosis, and enlargement of the coronary artery luminae.

Recently, assessment of CAC scoring by electron beam computed tomography (EBCT) or multi-detector CT (MDCT) has emerged as a potential noninvasive diagnostic technique for indirect detection of atherosclerotic burden in symptomatic patients.

1.2 The Technology and its Comparator(s)

Coronary artery calcification and detection by ultra-fast computed tomography

Coronary artery calcification (CAC) is part of the development of atherosclerosis. It is an active process that begins as early as the second decade of life and occurs exclusively in atherosclerotic arteries and is absent in the normal vessel wall.[12, 13]A close relationship has been confirmed both by histopathology and intravascular ultrasound between the extent of CAC and the atherosclerotic plaque burden seen in CAD, making calcium a potential marker for diseased arteries.[14, 15] Results from the St. Francis Heart Study showed that CAC scores in symptomatic patients, in both univariate and multivariate analyses, were independently predictive of CAD, surpassing the accuracy of historical cardiac risk factors.[1]Such studies, however, do not take into account treatment effects.

Early on, fluoroscopy was the modality of choice for detecting CAC. Around the 1990's, digital subtraction fluoroscopy and conventional computed tomography (CT) started being used and were reported to have greater sensitivity than conventional fluoroscopy in detecting CAC. More recently, ultra-fast computed tomographic scanning, which has the advantages of both rapid image acquisition, allowing the elimination of motion artifacts, and the high contrast and spatial resolution of computed tomography, has raised the question of whether it might be superior to other traditional methods for detecting and measuring CAC and thus a valuable tool in the diagnosis of CAD in symptomatic patients.

Coronary calcification is pervasive in patients with confirmed CAD and increases with age most markedly after age 50 in men and after age 60 in women.[16, 17] Increasing prevalence of coronary artery calcified plaque parallels the increasing prevalence of coronary atherosclerosis over the lifespan.[2] However, the presence of calcified coronary plaque is not strongly correlated with the extent of histopathologic stenosis.[18, 19] The

inner lining of both obstructed and non-obstructed vessels contains coronary artery calcified plaque; therefore, the detection of calcified plaque on cardiac CT is not specific to an obstructive lesion.[2] The absence of calcium in higher risk patients may not "rule out" the possibility of CAD. In a large study assessing the interrelation of CACS and inducible ischemia in patients with intermediate likelihood of CAD, Schenker, et al reported that 16% of patients with CACS = 0 had myocardial ischemia on provocative testing. [4]

The most common ultra-fast CT technologies employed for the detection of CAC are electron beam computed tomography (EBCT) and multi-detector computed tomography (MDCT). EBCT and MDCT for CAC measurement are completed within 10 to 15 minutes, requiring only a few seconds of scanning. Both methods use thin slice CT imaging and due to their extremely rapid scanning speed and use of ECG triggering and gating, have the advantage of reducing motion artifacts. Unlike most other cardiac CT studies for assessing heart morphology, function and coronary anatomy, CAC exams are performed without IV contrast medium and typically by prospective ECG-triggered sequential mode. Coronary artery calcium has a high intrinsic tissue contrast relative to non-calcified soft tissue so radiation dose can be reduced without diminishing the diagnostic value of the study.[20]The accuracy of ultra-fast CT is not limited by concurrent medications, the patient's ability to exercise, or baseline ECG abnormalities. Also, a large portion of the chest is imaged so non-cardiac causes of chest pain or other incidental findings may be identified or ruled out. (See the section on safety for more about incidental findings)

EBCT differs from conventional CTs (which generate x-ray images by mechanically rotating an x-ray tube) by electronically sweeping an electron-beam along a tungsten anode in a stationary tube. This design was developed in the 1980s in order to better image heart structures which are constantly in motion and performs a complete cycle of movement with each heart beat. EBCT testing is relatively inexpensive, costing around $420, which is comparable to a treadmill exercise test ($320), and significantly less than exercise echocardiography ($900) or dobutamine stress echocardiography ($1000) and coronary angiography (average $3000).[21] Still a relatively new technology, EBCT isn't use routinely to diagnose CAD because its accuracy hasn't been firmly established. Also, there is insufficient long term evidence to support the cost-effectiveness of CAC by EBCT over other conventional well established methods of testing.[1]

MDCT has recently emerged as another cardiac imaging modality that can be used to detect CAC. Although the current temporal resolution of MDCT is not as high as that for EBCT, MDCT as superior spatial resolution and unlike EBCT, can be used for the entire spectrum of routine clinical CT examinations.[20] MDCT differs from single detector-row helical or spiral CT scans in that it employs a two-dimensional array of detector elements versus a linear array, allowing multiple slices or sections (4 to 64) of the heart to be acquired simultaneously and with increased imaging speed. MDCT is a promising tool for CAC scoring but further studies are necessary to assess reproducibility and progression. One disadvantage of MDCT as compared with conventional CT is the higher degree of radiation exposure to the patient. In one study, radiation dose was

approximately 27% higher using MDCT versus single detector-row CT; organ dose was also higher with MDCT. [22]

Comparators

Exercise electrocardiography (ECG), stress echocardiogram, myocardial perfusion imaging (MPI), and thallium stress test are common, more established, noninvasive technologies used to diagnose and evaluate patients presenting with symptoms suspicious for CAD. Unlike calcium scoring or conventional angiography, which are anatomical tests, these noninvasive test allow assessment of ischemia and myocardial function.

An exercise treadmill test (ETT), also called an exercise stress test or an ECG treadmill test, evaluates the effect of exercise on the heart. ECG monitors the electrical activity of the heart and blood pressure readings are taken at various points throughout the test, measuring the heart's reaction to the body's increased need for oxygen. A stress test may be performed to determine the exercise capacity of the heart, the causes of chest pain, and to identify rhythm disturbances during exercise. The American College of Physicians recommends ETT as the first choice for primary care patients with symptomatic chronic stable angina or with a medium risk of CAD based on a set of common risk factors. [23] However, wide variability in sensitivity and specificity for exercise ETT has been reported across studies.[11]

Stress echocardiograms use ultrasound imaging to determine how the heart muscles respond to stress. A traditional stress echocardiogram requires patients to exercise on a treadmill or a stationary bike. For individuals who are unable to exercise, dobutamine can be used to increase the heart rate to a certain level, mimicking the effects of exercise. During both tests, blood pressure and heart rhythm (ECG) are monitored and ultrasound images are recorded that show areas of abnormal myocardial function which reflects reduced blood flow (ischemia) because of blocked arteries. A decrease in the left ventricular ejection fraction is an important indicator of possible damage to the heart or other problems with the heart valves or muscle, since the left ventricle performs the greatest amount of work pumping blood. Stress echocardiograms are simple and convenient to perform and do not require exposure to radiation.

Single photon emission computed tomography (SPECT) is a nuclear medicine imaging technique that employs a radioactive tracer which remains in the bloodstream and allows the visualization of blood flow to tissues and organs. When applied to the heart, it is often referred to as myocardial perfusion imaging (MPI) and is one of several types of cardiac stress tests used for the diagnosis of ischemic heart disease. MPI has been found to be comparable with other non-invasive stress tests for ischemic heart disease.[1] A thallium stress test is another nuclear imaging method that creates a picture of the heart by tracking how a radioactive material, thallium, moves through the vessels. These images show how well blood flows into the heart muscle after exercise or a medication-induced increase in heart rate. An abnormal thallium stress test is a possible indicator of a significant blockage of a coronary artery or damage to the heart muscle due to a prior heart attack. In direct-comparison studies of symptomatic persons, CAC detection has

been reported to be comparable to nuclear exercise testing in the detection of obstructive CAD. [1]

The current gold standard the anatomic assessment of CAD, and the main comparator for the purposes of this assessment, is coronary angiography, a minimally invasive procedure that uses x-ray imaging to visualize coronary anatomy. The procedure is performed by inserting a long, thin, flexible tube called a catheter into the body through the groin, arm, or neck, which is then threaded into the coronary arteries where a dye is injected into the bloodstream. While the dye is flowing through the coronary arteries, an x-ray machine rapidly takes a series of images, offering a detailed look at blood flow through the coronary arteries. Unlike cardiac CT-quantified calcium scoring, angiography detects obstructive CAD. Major complications are rare with coronary angiography but some potential risks may include bleeding, infection, and pain at the site of insertion, damage to the blood vessels, allergic reaction to the contrast dye, arrhythmias, blood clots, kidney damage, radiation exposure, heart attack, and stroke. Contrast-enhanced computed tomography (CTA) is a non-invasive method of visualizing the coronary arteries and might be useful as a substitute for coronary angiography in various clinical scenarios in which CAD us suspected.

Measures of coronary artery calcium

Agatston score – Named after Arthur Agatston, it is the most widely used method for quantifying CAC. The score is based on the area and the density of the calcified plaques as represented in Hounsfield units (HU), a quantitative measure of radiodensity used to evaluate CT scans. The pixel value of a CT scan is displayed according to the mean attenuation of the tissue that it corresponds to on a scale ranging from -1000 HU (air) to +1000 HU (bone), with water corresponding to 0 HU on the Hounsfield scale. In his landmark study in 1990, Agatston set the threshold for a calcified lesion at a density of 130 HU having an area of ≥ 1 mm in order to eliminate single pixels with a CT density > 130 units due to noise. A lesion score was then determined based on the maximal CT number: 1 = 130 to 199, 2 = 200 to 299, 3 = 300 to 399, and 4 ≥ 400 HU. A score for each region is calculated by multiplying the density score and the area and the total calcium score is then determined by adding up each of those score for all CT slices obtained.[24] [A test is considered positive if any calcification is detected within the coronary artery and negative if no calcifications are detectable. Agatston scores of < 10, 11-99, 100-400, and > 400 have been proposed to categorize individuals into groups having minimal, moderate, increased, or extensive amounts of calcification, respectively. The amount of calcium is related, to some extent, to the overall amount of atherosclerosis. [25]

Agatston CAC scores of 0, >100 and ≥ 400 are common cut-offs or thresholds seen throughout the literature to exclude or diagnose significant CAD. A score of 0 (no coronary artery calcified plaque) in symptomatic patients has demonstrated high negative predictive values, indicating it may be useful in excluding obstructive angiographic CAD.[2] A score of > 100 is considered a high calcium score and is consistent with a high risk of a cardiac event within the next 2 to 5 years (>2% annual risk).[13] Although there

are exceptions, a CAC score of ≥ 400 has been associated with an increased incidence of perfusion ischemia and obstructive CAD, making it a potential "high-risk" indicator for significant CAD.[2] It is important to remember that for diagnostic tests with continuous results such as the calcium score, the relative performance of sensitivity and specificity can be varied by changing the cut-off point defining positive or negative results. Which value is most important depends on one's objective and perspective.

Calcium volume score – A study by Callister in asymptomatic patients concluded that use of the calcium volume score showed better reproducibility than the traditional Agatston score.[26] However, partial volume effects may impair the score's accuracy, leading to overestimation of the calcium content and creating substantial variability between repeat trials.[27]

Calcium mass – Calcium mass has been shown to be very accurate with little variability but is not well validated and difficult to measure.[27] It is rarely used as a practical means of quantifying CAC.[28]

Safety
Radiation
The primary concern regarding the use of ultra-fast CT for calcium scoring is radiation exposure. To the extent that CACS may avoid the need for angiography, radiation exposure may be reduced. However, if CACS leads to additional testing, it may be increased. Even low levels of radiation are thought to increase the risk of cancer, though the extent by which this occurs is unclear.[2] In the American population, the collective dose received from medical uses of radiation was estimated to have increased by > 700% between 1980 and 2006. [29, 30]

When discussing the potential health risks related to radiation it is necessary to highlight the difference between radiation exposure and radiation dose. Radiation exposure is a measure of the quantity of ionization produced in air by photon irradiation. Radiation dose ("absorbed radiation dose") refers to the amount of radiation energy deposited in the human body as a result of exposure to ionization and is typically calculated from the exposure and from estimates of energy absorption per kilograms of body weight. [31]

The basic radiation dose parameter in CT is the computed tomography dose index (CTDI) and it is used to express the average dose delivered to the scan volume for a specific test.[32] Another important parameter is the effective dose (E), which is meant to reflect the risk of the biological effects of ionizing radiation and is useful in assessing and comparing the potential risk of a specific examination.[31] E is expressed in SI units of millisieverts (mSv) and corresponds to the amount of whole-body irradiation that would yield a biological risk equivalent to that of an irradiation to only a portion of the body.[33]It is not an exact indicator of the absolute risk of the biological effect on an individual but rather a rough estimate based on evolving knowledge and only applies to types of imaging studies, not to individual patients.[31] Though E should not be used for

epidemiological purposes, it is useful in comparisons of the biological risk of different medical procedures that use ionizing radiation.[33]

The effective dose values for coronary calcium CT have been reported to range from 1.0 to mSv to 12 mSv.[33] An AHA document reports that EBCT scanners have an estimated effective dose of 0.7 to 1 mSv in males and 0.9 to 1.3 mSv in females, and MDCT scanners have respective doses of 1 to 1.5 mSv and 1.1 to 1.9 mSv.[2] The estimated risk of fatal malignancy or death (per 1000 persons) resulting from CT calcium scoring (1 mSv) is 0.05, based on simulation.[33]

Factors that may influence the radiation dose include the CT scanner model, number and length of scans, scan mode, ECG triggering or gating, x-ray tube potential, tube current-time product, degree of overlap between adjacent CT slices (pitch), and patient size.[34] Recommendations reducing radiation are: to use a prospective ECG trigger as opposed to retrospective gating; to use EBCT scanners versus MDCT scanners; to dose radiation according to body size; to reduce the tube current-time product for small patients; and to not unnecessarily repeat examinations.[2]

Medical imaging is the largest controllable source of radiation exposure to the American public. [33] There is conflicting evidence regarding the risk of developing cancer at the levels and types of radiation associated with medical imagi ng and with the exception of mammography, there is currently no federal regulation of patient radiation dose.[31] Potential benefits and risks, including any related to not performing the test, should be carefully considered before ordering tests that will expose patients to ionizing radiation.

Incidental findings

Cardiac ultra-fast CT scanning for CAC includes images of portions of non-cardiac structures such as the lungs, bones, and upper abdomen. Pathologies unrelated to the heart or coronary arteries, both serious and benign, are sometimes identified incidentally when the entire scan is reviewed. The most common incidental finding is pulmonary nodules. Horton [35]analyzed 1326 patients with a mean age of 55 years undergoing EBCT for CAC screening and reported that lung nodules requiring clinical follow-up were seen in 65 (4.9%) patients. The prevalence of incidental findings in any organ system was 8%. In another study of 1000 middle-aged Army personnel, 23 (2.3%) were identified as having pulmonary nodules or other lung-related diseases, 50% of which were considered major, requiring subspecialty referral or potential invasive procedures. [36] The identification of potential pathology other than coronary calcium must be considered when evaluating the benefits and costs of cardiac scanning. [1]

1.3 Clinical Guidelines

No clear role for EBCT or MDCT calcium scoring in the diagnosis or prognosis of symptomatic patients has been clearly established in the available published literature. Furthermore, studies in symptomatic patients have not shown that clinical outcomes can be favorably altered by the used of CT-determined CAC for CAD.[1] One suggested

application in symptomatic patients may be to triage patients with suspected CAD. Proponents suggest that a CAC score of 0 can be used to rule out the likelihood of significant CAD.[1, 2] Another potential use of CAC is to determine the etiology of cardiomyopathy, specifically to differentiate ischemic from non-ischemic disease.[1]

Several clinical guidelines for the diagnosis and treatment of CAD were found on the National Guideline Clearinghouse (NGC) web-site, the primary repository for evidence-based clinical guidelines: http://www.guideline.gov.

American College of Cardiology Foundation Clinical Expert Consensus Task Force (ACCF/AHA)[1]

The ACCF/AHA 2007 Clinical Expert Consensus Document provided statements regarding the role of CAC by ultra-fast CT in clinical practice. The guidelines are summarized as follows:

- There is a lack of evidence from head-to-head studies comparing CAC measurement to alternative risk assessment techniques for moderate risk patients. At the current time, CAC measurement cannot be determined to be superior, or inferior, to other approaches for CAD risk assessment.
- No clear evidence is available indicating that additional non-invasive testing in patients with high (> 400) calcium scores will result in more appropriate selection of treatment over the currently recommended preventative medical therapies.
- Patients with atypical cardiac symptoms may benefit from CAC testing to help exclude the presence of obstructive CAD. Other competing modalities are available but most have not been compared directly to CAC.
- Available CAC data has come largely from studies in Caucasian, non-Hispanic men, thus discretion should be used in extrapolating current CAC data to women and ethnic minorities
- Current radiology guidelines should be used when determining need for follow-up testing of incidental findings on an ultra-fast CT study, such as was recently published to guide management of small pulmonary nodules. [MacMahon 2005]

American Heart Association [2]

In 2006, the AHA issued a scientific statement on the use of cardiac computed tomography which reviewed the efficacy of calcium scoring for determining prognosis and diagnosis.

Ratings of recommendations

The AHA often includes an assessment of quality of evidence underlying the recommendation and the benefit versus risk using the following scoring system:

Evidence Level

Level A: Multiple randomized clinical trials or meta-analysis

Level B: Single randomized clinical trial or observational data (case control, longitudinal data)
Level C: Case reports, expert opinion, or current clinical practice

Benefit versus risk
Class I: Benefit >>> risk; procedure or treatment SHOULD be performed (i.e. is recommended, indicated, useful/effective/beneficial)
Class IIa: Benefit >> risk; procedure or treatment IS REASONABLE to perform
Class IIb: Benefit < risk, procedure or treatment MAY BE CONSIDERED
Class III: Risk outweighs the benefit; procedure SHOULD NOT be performed

Conflicting evidence and/or a divergence of opinion regarding its usefulness was found for the following indications:

- Symptomatic patients with chest pain with equivocal or normal electrocardiograms and negative cardiac enzymes (Class IIb, Level of Evidence: B)
- Determining the etiology of cardiomyopathy (Class IIb, Level of Evidence: B)
- Symptomatic patients in the setting of ambiguous stress tests (Class IIb, Level of Evidence: B)
- Asymptomatic patients with intermediate risk of CAD (Class IIb, Level of Evidence: B). May be useful to refine clinical risk prediction and to select patients for more aggressive target values for lipid-lowering therapies.

Indications in which CAC scoring was deemed not useful or possible harmful:

- Low-risk or high-risk asymptomatic patients (Class III, Level of Evidence: B)
- Establishing the presence of obstructive disease for revascularization in asymptomatic persons (Class III, Level of Evidence: C)
- Serial imaging for assessment of progression of coronary calcification (Class III, Level of Evidence: C)
- Hybrid nuclear and CT imaging to assess cardiovascular risk or presence of obstructive disease (Class III, Level of Evidence C)

Furthermore, the report stated that despite growing evidence that calcium scores are an independent predictor of CAD studies have not demonstrated improved clinical outcomes as a results of calcium score screening.

American Heart Association

According to a 2009 scientific advisory from the AHA, the following are the minimum requirements which should be met in scanning for coronary artery calcium (CAC) [33, 37]:

- Use of an EBCT scanner or a 4-level (or greater) MDCT scanner
- Cardiac gating
- Prospective triggering for reducing radiation exposure
- A gantry rotation of at least 500 ms

- Reconstructed slice thickness of 2.5 to 3 mm to minimize radiation in asymptomatic persons (and to provide consistency with established results)
- Early to mid-diastolic gating
- Equipment or nuclear material in cardiac imaging should be appropriately utilized to maintain patient doses as low as reasonable achievable (ALARA) but consistent with obtaining the desired medical information

American College of Radiology (ACR) Appropriateness Criteria 2008
For assessment of chronic chest pain in patients with low to intermediate probability of CAD:

- CT coronary calcium scoring received a rating of 3 (1 = least appropriate, 9 = most appropriate)
- A score of zero may be useful in excluding cardiac etiology
- Relative radiation level is considered to be medium based on the following:

Relative Radiation Level Designations

Relative Radiation Level	Effective Dose Estimate Range
None	0
Minimal	< 0.1 mSv
Low	0.1-1 mSv
Medium	1-10 mSv
High	10-100 mSv

ACC/AHA expert consensus document on EBCT for the diagnosis and prognosis of CAD [2, 13]
According to a statement by the ACC/AHA in 2000, the following are interpretations and recommendations for cardiac CT scanning and CAC scoring:

- A negative test (score = 0) makes the presence of atherosclerotic plaque, including unstable or vulnerable plaque, highly unlikely.
- A negative test makes the presence of significant luminal obstructive disease highly unlikely.
- A negative test is consistent with a low risk (0.1% per year) of a cardiovascular event in the next 2 to 5 years.
- A positive test (CAC > 0) confirms the presence of a coronary atherosclerotic plaque.
- The greater the amount of coronary calcium, the greater the atherosclerotic burden in men and women, irrespective of age.
- The total amount of coronary calcium correlates best with the total amount of atherosclerotic plaque, although the true atherosclerotic burden is underestimated.
- A high calcium score (Agatston score >100) is consistent with a high risk of a cardiac event within the next 2 to 5 years (>2% annual risk).
- CAC measurement can improve risk predication in conventional intermediate-risk patients, and CAC plaque scanning should be considered in individuals at intermediate risk for a coronary event (1.0% per year to 2.0% per year) for clinical decision-making with regard to refinement of risk assessment.
- Decisions for further testing beyond assistance in risk stratification in patients with a positive CAC score cannot be made on the basis of coronary calcium scores

alone, as calcium score, correlates poorly with stenosis severity in a given individual and should be based upon clinical history and other conventional clinical criteria.

1.4 Previous Technology Assessments

Thirteen Health Technology Assessments were found and reviewed, ten of which either performed CAC scoring analyses in asymptomatic patient populations or using computed tomographic angiography (CTA), and were therefore outside the scope of this report and were excluded from assessment.

- Agency for Healthcare Research and Quality (AHRQ), October 3, 2006[38]
- Ontario Health Technology Advisory Committee (OHTAC), May 2007[39]
- BlueCross BlueShield Technology Assessment, August 2006[40]
- Berry E et al. (NHS R&D Health Technology Programme), October 1999[41]
- Canadian Agency for Drugs and Technologies in Health (CADTH), August 2006[42]
- National Horizon Scanning Centre, December 2006[43]
- Belgian Health Care Knowledge Centre (KCE), 2008[44]
- Mowat G, et al. (NHS R&D Health Technology Programme), May 2008
- New Zealand Health Technology Assessment (NZHTA) Evidence Tables, February 3, 2003[45]
- Waugh N et al (NHS R&D Health Technology Programme), October 2006[46]

Three Health Technology Assessments provided limited data or information and are briefly described below:

- The AHRQ did a review of various non-invasive technologies for diagnosing CAD in symptomatic women.[47] The overall accuracy of CT calcium scoring was found to be low in both men and women, cut-offs of both > 0 and > 100. A calcium score of 0, however, had a high sensitivity and a low negative likelihood ratio is low, indicating that a calcium score of 0 might be useful to rule out CHD in both women and men.
- The California Technology Assessment Forum (CTAF) investigated the utility of CAC measurement in cardiovascular disease. [48] The CTAF reviewer recommendation was as follows:

  > As a diagnostic test in patients with symptoms suggestive of CAD does not meet technology assessment criteria 3, 4, or 5 for safety, effectiveness, and improvement in health outcomes.

  The report then states that, following clarification of data and testimony from invited experts, the CTAF panel accepted the following recommendation:

  > As a diagnostic test in patients with symptoms suggestive of CAD (i.e. chest pain) EBCT calcium scoring was determined to be a useful

technology in the prediction of those patients who will have underlying coronary disease.

- The Institute for Clinical Systems Improvement (ICSI) Technology Assessment Committee reported that for the diagnosis of obstructive CAD in symptomatic patients, EBCT or helical CT CAC score is a stronger independent predictor than conventional risk factors.[49] Direct comparisons of EBCT with other non-invasive tests for diagnosis are lacking, however.

1.5 Medicare and Representative Private Insurer Coverage Policies

The table below provides a summary of payer policies related to coronary artery calcium scoring.

Table 3. Summary of payer policies

Payer (year)	Evidence base available	Policy	Rationale
Centers for Medicare & Medicaid Services (CMS):[50, 51] **Pub 100-3 National Coverage Determinations: Section 220.1, Version 2** **NCA tracking #: CAG-00385N** **(2008)**	NR	• No national coverage determination for CTA is appropriate at this time (March 12, 2008); use of CTA to diagnose coronary artery disease will remain at local contractor discretion • No mention of calcium scoring found	• After examining the medical evidence available for CTA in the visualization of coronary arteries, no national coverage determination is appropriate at this time (March 12, 2008) • There is limited evidence regarding the test performance of non-invasive imaging tests for identifying, quantifying, or otherwise characterizing coronary artery stenosis
CMS Regional Coverage **LCD ID number L23654 (2008)** **And** **Article (A45280) (Washington, Alaska), administered by Noridian Administrative Services (2007)**	NR	• Demonstration and/or quantification of the presence of coronary calcification in either asymptomatic or symptomatic patients with or without signs of atherosclerotic heart disease has not been shown to improve outcomes and is not covered. Until such time as there may be more evidence of medical necessity, Medicare will not pay for the quantitative evaluation of coronary calcium by MDCT, CTCA, EBCT or other technology. • Cardiac MDCT for coronary calcium scoring is not covered • CPT codes: 0144T	• Lack of evidence of the medical necessity for quantitative evaluation of coronary calcium

Aetna Clinical Policy Bulletin number 0228 (2009)[52]	Unable to determine	• Calcium scoring is considered medically necessary for diagnostic cardiac CT angiography to assess whether an adequate image of the coronary arteries can be obtained
		
		  • CPT codes: 0145T, 0146T, 0147T, 0148T, 0149T, 1050T, +0151T
		
		  • ICD-9 codes: 424.3, 446.1,
			• NR

		745.2, 747.40-747.42, 759.82 • Calcium scoring (e.g., with ultrafast EBCT, spiral (helical) CT, and multislice CT) is considered experimental and investigational for all other indications • CPT codes: 0144T • HCPCS codes: S8092	• Definitive value of calcium scoring for assessing coronary heart disease risk has not been established in the peer-reviewed published medical literature
Cigna HealthCare Coverage Position number 0009 (2008)[53]	Symptomatic patients 5 studies, N = 4821 Asymptomatic patients 14 studies, N = 90,253	• EBCT, spiral CT, or MDCT for the detection and/or quantification of coronary artery calcification is not covered • Use of these technologies for screening, diagnosis, or management of coronary artery disease is considered experimental, investigational, or unproven • CPT codes: 0144T • HCPCS codes: S8092 • ICD-9 codes: 414.01	• There is insufficient evidence to support the use of MDCT and EBCT scanning for coronary calcium in symptomatic or asymptomatic populations
Civilian Health and Medical Program of the Department of Veterans Affairs (CHAMPVA)[54] **Policy Manual, Benefits, Chapter 2, Section 26.3 (2008)**	NR	• Helical CT for assessment of coronary calcification is not covered • CT or EBCT to screen asymptomatic individuals for CAD is not covered	• Cardiac CT including image post processing and quantitative evaluation of coronary artery calcium is unproven • CTA of coronary arteries (including native and anomalous coronary arteries, coronary bypass grafts) with quantitative evaluation of coronary calcium is unproven
Regence of Oregon and Utah; Regence of Idaho and select counties of Washington Medical Policy number 6 (2008) [55]	2006 AHA scientific statement 1999 ACC/AHA Expert Consensus Document	• Use of CT to detect and quantify coronary artery calcification is considered investigational • CPT codes: 0144T, 1047T, 1049T • HCPCS codes: S8092	• Published studies do not establish a clear role for EBCT in coronary risk stratification is asymptomatic or symptomatic patients • No studies have show that clinical outcomes can be favorable altered by the use of CT-based determination of coronary artery calcification in screening for coronary artery disease

BlueCross BlueShield (BCBS) of North Carolina Corporate Medical Policy number RAD5050 (2008)[56]	NR	• CT to detect coronary artery calcification is not covered
		• Use of CT (e.g., EBCT, spiral or helical CT, multislice or multi-detector CT) to detect coronary artery calcification is considered investigational for all services including:
		  • screening examination for asymptomatic patients
		  • diagnostic study in symptomatic patients
		  • assessment of coronary artery
			• NR for diagnostic purposes
			• No clear role has been established in the literature for EBCT in coronary disease risk stratification in asymptomatic patients
			• No studies have shown that clinical outcomes can be favorably altered by the use of screening EBCT

		bypass graft patency • measurement of cardiac perfusion • CPT codes: 0144T, 0147T, 0149T • HCPCS codes: S8092	
United Health Care (July 2009)	NR, appears to use information from ACC/AHA statements and guidelines	• Calcium scoring as a triage tool for symptomatic patients to rule out obstructive disease and avoid invasive procedures is covered • Calcium scoring for routine screening is not covered but risk stratification in asymptomatic patients with moderate risk based on the Framingham Score is covered	• States that CACS for triage is proven

ACC: American College of Cardiology; AHA: American Heart Association; CPT: current procedural technology; CT: computed tomography; CTA: computed tomographic angiography; EBCT: electron beam computed tomography; HCPCS: healthcare common procedure coding system; ICD-9: international classification of diseases, 9th addition; MDCT: multidetector computed tomography; NR: not reported

- **Medicare (National Coverage Determination)**
  The most current information available since March 2008 indicates that The Centers for Medicare and Medicaid Services (CMS) find that no national coverage determination for computed tomographic angiography (CTA) is appropriate at this time. No mention of coronary artery calcium scoring could be found.

- **Medicare (Regional Coverage Determination)**
  The local regional CMS had determined that there is a lack of evidence of the medical necessity for quantitative evaluation of coronary artery calcium. Coronary artery calcium scoring by multidetector computed tomography (MDCT) is not a covered service. LCD ID number L23654, states: "Demonstration and/or quantification of the presence of coronary calcification in either asymptomatic or symptomatic patients with or without signs of atherosclerotic heart disease has not been shown to improve outcomes and is not covered. Until such time as there may be more evidence of medical necessity, Medicare will not pay for the quantitative evaluation of coronary calcium by MDCT, CTCA, EBCT or other technology."

- **Aetna**
  Aetna considers calcium scoring medically necessary for diagnostic cardiac CT to assess whether an adequate image of the coronary arteries can be obtained. Calcium scoring with electron beam CT, spiral or helical CT, and multislice CT is considered experimental and investigational for all other indications. To date, the peer-reviewed published medical literature has produced conclusive evidence of value of calcium scoring for assessing coronary artery disease (CAD) risk.

- **Cigna**

Cigna considers electron beam CT, spiral or helical CT, or multidetector CT for screening, diagnosis, or management of coronary artery disease experimental, investigational, or unproven. The detection and/or quantification of coronary artery calcification using these technologies are not a covered service.

- **Department of Veterans Affairs (VA)**
  The medical program for the VA does not cover the use of helical CT scanning for the assessment of coronary calcification. Cardiac CT scanning including image post processing and quantitative evaluation of coronary artery calcium is considered unproven. Likewise, CT or EBCT to screen asymptomatic individuals for CAD and CT angiography with quantitative evaluation of coronary calcium are not covered due to lack of evidence supporting their validity.

- **Regence (Regional Medical Policy)**
  Regence's local medical policy for the states of Oregon, Idaho, Utah, and select counties of Washington considers the use of CT to detect and quantify coronary artery calcification investigational. No clear role for EBCT in coronary risk stratification in asymptomatic or symptomatic patients has been established in the available published literature. Furthermore, no studies have show that clinical outcomes can be favorably altered by the use of CT-based determination of coronary artery calcification in screening for coronary artery disease.

- **BlueCross BlueShield of North Carolina (Corporate Policy)**
  BlueCross BlueShield does not cover the use of CT to detect coronary artery calcification and is considered investigational for all services.

1.6 Washington State Data

Data from three Washington State Agencies were provided by the Health Technology Assessment Program. HTA coordinates the collection of any relevant agency utilization data. Coronary Artery Calcium Scoring (CACS) is a selected topic. CACS uses a CT to check for the buildup of calcium in plaque on the coronary arteries. This test identifies and quantifies a marker of coronary disease (plaque) and advocates believe it detects earlier stage CAD (before it becomes clinically apparent) that can be intervened on through a combination of non-invasive (lifestyle and medication) or invasive (angiography, stent, CABG) approaches.

Estimates for costs and utilization from the Uniform Medical Plan (UMP) and Washington State's Medicaid Program (DSHS) are presented below in Table A. They provide an estimate of base costs and may not include all costs for Coronary Artery Calcium Scoring (CACS). Information on relevant procedure codes is included after the result tables.

Current State Agency Medical Policy

Medicaid: CCTA is currently a covered service and requires "preauthorization" by Medicaid clinical utilization review consultants.

Uniform Medical Plan: CCTA is currently a covered service only by Exception, subject to preauthorization review. In most cases it was deemed **"investigational"** by UMP medical consultants. According to UMP's Summary of Benefits, a service or supply is considered experimental or investigational if it is under continued scientific testing and research concerning safety, toxicity, or efficacy and is unsupported by prevailing opinion among medical experts (as expressed in peer-reviewed literature) as safe, effective, and appropriate for use outside the research setting. Providers may request an exception through the UMP medical review staff.

Labor and Industries: This service is not generally within the scope of services covered because heart disease and diagnosis is not typically related to a work place injury. If requested and within scope of services, it would be considered under WAC 296-20-01002 which outlines that in no case shall services which are inappropriate to the accepted decision or which present hazards in excess of the expected medical benefits be considered proper and necessary. Services that are controversial, obsolete, investigational or experimental are presumed to not be proper and necessary. Providers may request an exception through the medical director.

Table A: Claims by Year
UMP & Medicaid

CPT CODE	2005	2006	2007	2008	Total
0144T (CT, heart, w/o contrast, with eval of coronary calcium)	0	12	27	53	**92**
0147T (CT angiography of coronary arteries with eval of coronary calcium)	0	13	25	19	**57**
0149T (Cardiac structure and morphology and CT angiography with eval of coronary calcium)	0	0	17	11	**28**
Total	**0**	**25**	**69**	**83**	**177**

Paid Claims by Year
UMP & Medicaid

CPT CODE	2005	2006	2007	2008	Total
0144T (CT, heart, w/o contrast, with eval of coronary calcium)	0	1	3	23	**27**
0147T (CT angiography of coronary arteries with eval of coronary calcium)	0	8	20	11	**39**
0149T (Cardiac structure and morphology and CT angiography with eval of coronary calcium)	0	0	13	9	**22**
Total	**0**	**9**	**36**	**43**	**88**

Total Payments* by Procedure by Year
UMP & Medicaid | 2005-2008

CPT Code	2005	2006	2007	2008	Total
0144T	$0	$149	$1,173	$3,031	**$4,353**
0147T	$0	$1,645	$4,043	$4,593	**$10,281**
0149T	$0	$ 0	$7,331	$4,648	**$11,979**
Total	**$0**	**$1,794**	**$12,547**	**$12,272**	**$26,613**

**Payments include professional and facility fees.*

Average Payments* by Procedure by Year

UMP & Medicaid | 2005-2008

CPT Code	2005	2006	2007	2008	Total
0144T	$0	$149	$391	$132	**$161**
0147T	$0	$206	$202	$418	**$264**
0149T	$0	$ 0	$564	$516	**$577**
Total	**$0**	**$199**	**$349**	**$285**	**$302**

Procedure Codes

CPT Codes

0144T – non contrast cardiac CT with eval of calcium scoring
0147T – cardiac CT angiography with calcium scoring
0149T – cardiac structure and morphology and cardiac CT angiography with calcium scoring

Cardiac Computed Tomography and Coronary Computed Tomographic Angiography (APCs 0282 and 0383)
For CY 2008, CMS will assign the cardiac computed tomography (CCT) and coronary computed tomographic angiography (CCTA) procedures to two new clinical APCs, specifically new clinical APC 0383 (cardiac computed tomographic imaging) and APC 0282 (Miscellaneous Computed Axial Tomography). The median cost of approximately $314 for APC 0383 was based entirely on claims data for CPT® Category III codes 0145T, 0146T, 0147T, 0148T, 0149T, and 0150T that described CCT and CCTA services, a clinically homogeneous grouping of services. In addition, the individual median costs of these services ranged from a low of approximately $277 to a high of $437, reflecting their hospital resource similarity as well. CMS proposed to reassign the other two CCT CPT® codes, specifically CPT® codes 0144T and 0151T, to APC 0282. The inclusion of these two codes in APC 0282 resulted in a CY 2008 APC median cost of about $105.

2. The Evidence

2.1 Systematic Literature Review

Objectives

The primary aim of this assessment is to systematically review, critically appraise and analyze research evidence describing the performance and safety of CACS as a *diagnostic* test for evaluation of CAD in patients in whom CAD is suspected. Available information on the economic impact of this will also be summarized and critically appraised.

2.2 Methods

Inclusion and exclusion criteria: For Key Questions:

- Retrieval and assessment of studies focused on those of the highest methodological quality.
- The focus of this HTA is on diagnostic application of CACS (as a stand-alone test) in *symptomatic* patients as opposed to screening of asymptomatic patients.

Table 4. Summary of inclusion and exclusion criteria are in the table:

Study Component	Inclusion	Exclusion
Participants	• Symptomatic patients with suspected CAD who have not had revascularization (i.e. CABG or stent) • Consecutively enrolled patients • Studies relate to triage of chest pain patients (emergency department)	• Asymptomatic patients and those judged to be at low risk for CAD • Patients who have had previous revascularizaton (CABG, PTCA or stenting) • Studies of serial assessment of CAC • Evaluation of other cardiac diseases (e.g. valvular disease etiology of cardiomyopathy)
Intervention	• Coronary artery calcium scoring (CACS) using computed tomography (EBCT, MDCT, spiral/helical CT, multi-slice CT)	• MRI
Reference Standard:	• Coronary artery angiography is reference standard • Studies comparing CACS with other diagnostic tests will only be briefly described for context.	•
Outcomes	• Test performance/accuracy parameters (sensitivity, specificity, PPV, PVN, and reliability) • Death, myocardial infarction, patient-reported outcomes • Economic measures (e.g. ICERs, cost per correct diagnosis)	• studies focused on "per-vessel" or "per-segment" analysis without per patient findings
Study Design	• Prospective studies directly comparing CACS with the reference standard or other comparators (except reliability studies) will be sought. Retrospective studies will be considered if there are insufficient prospective studies. • CT and coronary angiography carried out within 3 months of each other • For Key Question 5, only formal economic studies will be considered	• Case series • Case reports • Studies that do not directly compare CACS using CT with conventional angiography
Publication	• Studies published in English in peer reviewed journals or publically available FDA reports • For Key Question 5 o Full formal economic analyses (e.g. cost-utility studies) published in English in a peer-	• Abstracts, editorials, letters • Duplicate publications of the same study which do not report on different outcomes • Single reports from multicenter trials

reviewed journal published after those represented in previous HTAs.	• White papers • Narrative reviews • Articles identified as preliminary reports when results are published in later versions • Incomplete economic evaluations such as costing studies

Data sources and search strategy

The clinical studies included in this report were identified using the algorithm shown in Figure 1 below. The search took place in four stages. The first stage of the study selection process consisted of a comprehensive literature search using electronic means and hand searching. We then screened all possible relevant articles using titles and abstracts in stage two. This was done by two individuals independently. Those articles that met a set of *a priori* retrieval criteria based on the criteria above were included. Any disagreement between screeners that were unresolved resulted in the article being included for the next stage. Stage three involved retrieval of the full text articles remaining. The final stage of the study selection algorithm consisted of the selection of those studies using a set of a priori inclusion criteria, again, by two independent investigators. Those articles selected form the evidence base for this report.

Figure 1. Algorithm for article selection

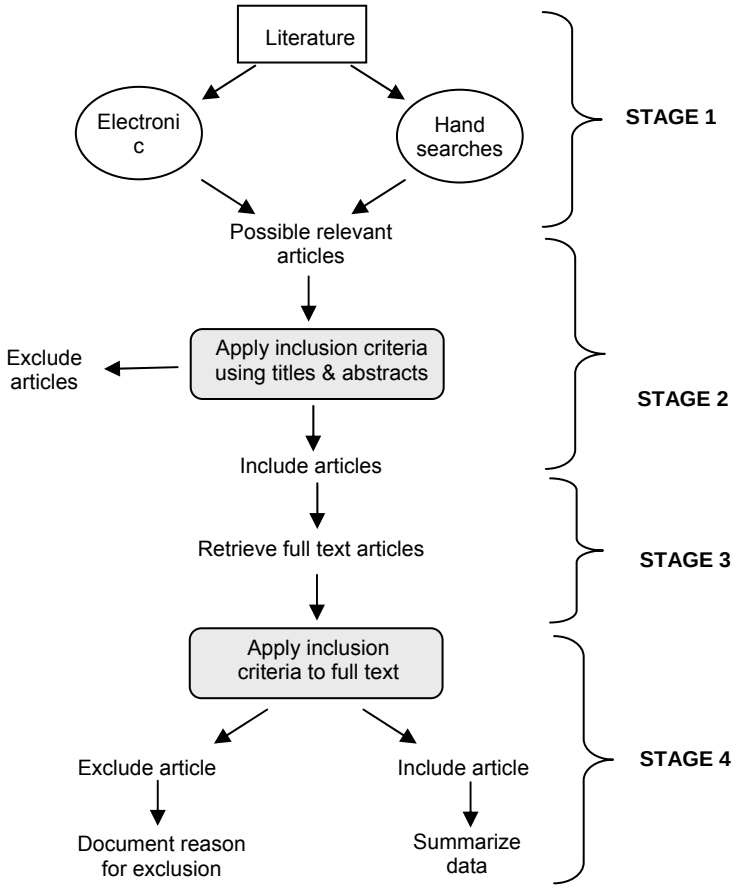
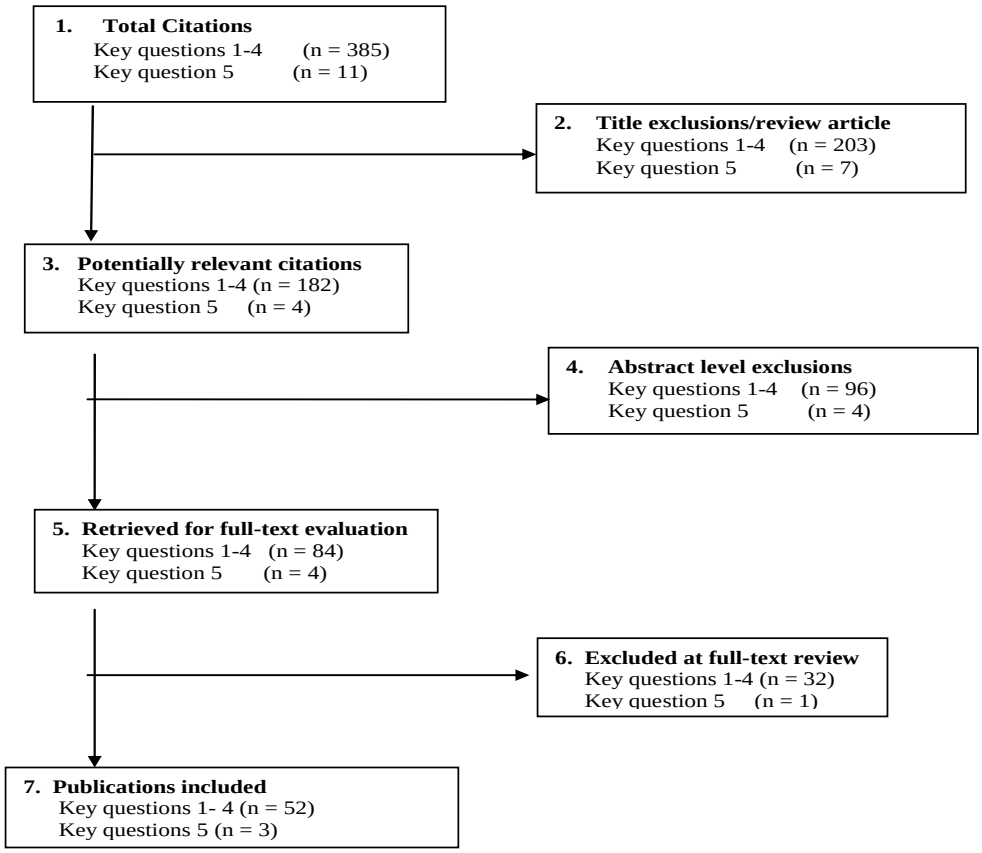


Figure 2. Flow chart showing results of literature search



Categorization of studies and outcomes

Data extraction

For the highest quality (level of evidence I – III, please see below) validation studies, population characteristics, study design, inclusion criteria, diagnostic test characteristics, prevalence of CAD and outcomes were abstracted. Where possible, true positive, true negative, false positive and false negative results were abstracted (or calculated) for meta-analysis based on data provided in the reports.

Study quality assessment: Level of evidence (LoE) evaluation

Details of the Level of Evidence (LoE) methodology are found in Appendix C. Each validation, reliability or economic study chosen for inclusion was given a LoE rating based on the quality criteria described below. Abstraction guidelines were used to determine the LoE for each study included in this assessment. The methodological quality of studies was independently assessed by two reviewers and discrepancies were resolved by discussion.

The method used by Spectrum Research, Inc. (SRI) for assessing the quality of evidence of individual studies as well as the overall quality of evidence incorporates aspects of the

rating scheme developed by the Oxford Centre for Evidence-based Medicine, [57] precepts outlined by the Grades of Recommendation Assessment, Development and Evaluation (GRADE) Working Group, [58] and recommendations made by the Agency for Healthcare Research and Quality (AHRQ).[59] Details are found in Appendix C.

For validation (accuracy) studies, the following criteria were used to assess LoE for individual studies:

Definitions of the different levels of evidence for diagnostic test accuracy/validity studies.

Level	Study type	Criteria
I	Good quality prospective study	• Broad spectrum of persons with the expected condition • Appropriate reference standard used • Adequate description of test and reference for replication • Blinded comparison of tests with appropriate reference standard • Reference standard performed independently of diagnostic test
II	Moderate quality prospective study	• Violation of any one of the criteria for a good quality prospective study (LoE I)
	Good quality retrospective study	• Broad spectrum of persons with the expected condition • Appropriate reference standard used • Adequate description of test and reference for replication • Blinded comparison of tests with appropriate reference standard • Reference standard performed independently of diagnostic test
III	Poor quality prospective study	• Violation of any two or more of the criteria for a good quality prospective study (LoE I)
	Moderate quality retrospective study	• Violation of any one of the criteria for a good quality retrospective study (LoE II)
IV	Poor quality retrospective study	• Violation of any two or more of the criteria for a good quality retrospective study (LoE II)
	Case-Control Study	

For reliability studies, the following criteria were used to determine the LoE for individual studies:

Definitions of the different levels of evidence for reliability studies

Level	Study type	Criteria

I	Good quality study	• Broad spectrum of persons with the expected condition • Adequate description of methods for replication • Blinded performance of tests, measurements or interpretation • Second test/interpretation performed independently of the first
II	Moderate quality	• Violation of any one of the criteria for a good quality study
III	Poor quality study	• Violation of any two of the criteria
IV	Very poor quality study	• Violation of all three of the criteria

There is no universally accepted, standardized approach to critical appraisal of economic evaluation studies. The criteria described in the Quality of Health Economic Studies (QHES) tool[60] provided a basis for the critical appraisal of included economic studies and was augmented with the application of epidemiologic appraisal precepts (see Appendix C). The QHES employs widely accepted criteria for appraisal, such as choice and quality of cost and outcomes measures, transparency of model and presentation, use of incremental analysis, uncertainty analysis, and discussion of limitations and funding source and was primarily used to facilitate description of primary strengths and limitations of the studies. A weighted global score can be obtained based on these measures with a possible range of scores from 0 (worst) to 100 (best), theoretically providing a common metric to compare study quality. This tool and the weighted score have not yet undergone extensive evaluation for broad use but provide a valuable starting point for critique.

Two individuals critically appraised each study independently using the QHES. Discrepancies were resolved by discussion to arrive at a final appraisal. In addition, elements of critical appraisal consistent with epidemiologic principles and evaluation of bias (e.g., selection bias) were applied. Evaluation of the overall strength of evidence across studies for specific key questions, considers the quality and quantity of available studies as well as the consistency of study estimates.

Overall Strength of Evidence (SoE)

SoE	Description	Further Research Impact	Domain Criterion Met		
			Quality	Quantity	Consistency
1	**High**	Very unlikely to change confidence in effect estimate	+	+	+
2	**Moderate**	Likely to have an important impact on confidence in estimate and *may* change the estimate	+	-	+
			+	+	-
3	**Low**	Very likely to have an important impact on confidence in estimate and *likely* to change the estimate	+	-	-
			-	+	+
4	**Very Low**	Any effect estimate is uncertain	-	+	-
			-	-	+
			-	-	-

Data analysis
Meta-analysis was conducted on the primary validation parameters when data from three or more LoE I-III studies were available. Estimates from LoE I and II studies were pooled separately from estimates for LoE III studies, providing some sensitivity analysis related to study quality. Pooled estimates of sensitivity, specificity, positive predictive value and 1-negative predictive value were calculated.

There is no consensus on thresholds (cut offs) for normal or abnormal CACS scores. One meta analysis and two documents from the American College of Cardiology (ACC) have suggested that a cut off score of 0 might be assumed to be associated with a low risk (≤ 5%) of CAD and potentially used to rule out significant obstructive CAD while a score of ≥ 400 was assumed to be associated with a 90% risk of CAD.[1, 2, 61] Following consultation with a clinical expert, these two primary cut points were chosen for primary meta-analysis based on the premise that a score of 0 might be used to "rule out" significant CAD and a score of ≥ 400 might signal the need for additional evaluation. Thresholds and methods of determining optimal cut points for scores varied across studies and could not be pooled. Most studies used an angiographic cut off of ≥ 50% decrease in luminal dimension for determination of obstructive CAD.

Meta-analysis using the above thresholds was done if data for true negative, true positive, false negative and false positive could be abstracted or calculated. Meta-analysis was performed using MetaDiSc software version 1.4[62] and summary ROC curves were calculated using standard methods. [63]

2.3 Quality of literature available
From a list of 186 potentially relevant study citations from electronic database searches, no randomized controlled trials were found. Multiple reports of the same study and/or of overlapping populations were found. Reports representing the most complete data were retained.

Five meta-analyses were found,[1, 13, 61, 64, 65] two of which were contained within ACC/AHA guideline documents.[1, 13] All of these analyses included individual studies that varied widely with respect to quality, based on Spectrum's LoE determination and only one provided some evaluation of individual study quality.[61] One analysis appears to have combined studies with different population characteristics (i.e. those with suspected CAD, chronic dialysis patients, those with non-ischemic cardiomyopathy) but did describe potential sources of heterogeneity. [65] The meta-analysis reported by Heijenbrok-Kal included 21 EBCT studies but it was unclear whether these studies were in symptomatic or asymptomatic populations.[64] Since the quality of studies varied in these meta-analyses and several appeared to have included populations not relevant to this technology assessment, information from them is provided where appropriate to provide additional context only.

Accuracy (validation) studies
A total of 30 primary studies of accuracy and validity comparing CACS with CCA were identified. Of these, 11 studies[66-76] were classified as LoE I or II, 8 as LoE III[77-84] and 11

as LoE IV[14, 85-91] as listed below. Two studies included evaluation of CACS in diabetic populations.[70, 82] Not all LoE I –III studies had data that could be abstracted for meta-analysis.

Blinded interpretation of the test and reference standard results and independent performance of the test and referent (i.e. the results of one should not influence whether the other is performed) are considered to be of primary importance in decreasing study bias. It was not always clear from reports that these criteria were met. A brief summary of the level of evidence (LoE) determination for these studies is found below.

LoE I	**LoE II**	**LoE III**	**LoE IV**
Leschka 2008	Lau 2005	Hosoi 2002 (DM vs. not)	Konieczynska (2006)
Kajinami 1995	Nixdorf 2008	Shavelle 2000	Haberl 2005
	Becker 2007	Budoff 1996†	Rumberger, Sheedy 1997
	Knez 2004	Tannenbaum 1989	Yao 2004
	Lamont 2002	Herzog 2004	Shivastava 2003
	Haberl 2001	Fallavollita 1994	Broderick 1996
	Leber 2001	Chen 2001	Bielak 2000
	Budoff,2002	Guerici 1998	Yao 2000
	Kwok 2000		Seese 1997
	Khaleeli (2001) †		Baumgart 1997
			Bielak 1994

†Khaleeli focuses on CACS in diabetic patients but compares findings with a non-diabetic cohort previously reported in Budoff 1996.

Table 5. Level of evidence (LoE) summary for LoE I/II validation studies

METHODOLOGICAL PRINCIPLE	Leschka (2008)	Nixdorff (2008)	Becker (2007)	Lau (2005)	Knez (2004)	Budoff (2002)	Lamont (2002)	Haberl (2001)	Leber (2001)	Kwok (2000)	Kajimani (1995)
Study Design											
Prospective cohort design	√	√		√			√				√
Retrospective cohort design			√		√	√		√	√	√	
Case-control design											
Broad spectrum of patients with expected condition	√		√		√			√	√	√	√
Appropriate reference standard	√	√	√	√	√	√	√	√	√	√	√
Adequate description of test and reference for replication	√	√	√	√	√	√	√	√	√		√
Blinded comparison with appropriate reference	√	√	√	√	√	√	√	√	√	√	√
Reference standard performed independently of test	√	√	√	√	√	√	√	√	√	√	√
Evidence Level	**I**	**II**	**II**	**II**	**II**	**II**	**II**	**II**	**II**	**II**	**I**

* Blank box indicates criterion not met or could not be determined or information not reported by author

Conventional coronary angiography (CCA) was considered to be the best primary reference test for comparing accuracy of CACS. Since symptomatic patients are most likely to be referred by a cardiologist for angiography as a definitive test based on a clinical suspicion of CAD, there is the potential for referral/verification bias inherent in all studies. Verification bias occurs when results of the index test (CACS here) are used implicitly or explicitly to determine if a subject should have the reference test procedure (coronary arteriography). If so, estimates of sensitivity tend to be biased upward while estimates of specificity tend to be biased downward. However, verification bias does not affect positive and negative predictive values.[92]

In a large number of studies, it was unclear whether CCA and CACS were performed independently of each other. Few authors specifically stated that the decision to perform one was independent of the decision to perform the other. Where independence could be reasonably inferred, credit was given for this, however, the potential for bias may still be present to the extent that there the results of one test may have influenced the decision to perform the other. To the extent to which there is not independence and blinded interpretation, there is the potential for referral/verification bias which may lead to overestimation of test accuracy.

The definition of a "broad spectrum" of patients who are most likely to receive the test now or in the future would be a reflection of the fact that this is primarily a referral population with suspected CAD

While some studies explicitly stated that study design and data collection were prospective, for the largest proportion of studies, it was not clear if the study was prospective or retrospective and no credit for this criterion could be assigned. Most were presumably retrospective analyses even though some data may have been collected prospectively.

Author reporting of blinded interpretation of both tests was not consistent. Several studies indicated that blinded interpretation of one study was done, but did not report that the interpretation of the other study was done in a blinded fashion.

Two of the validation studies[69, 83] also provided data validating other non-invasive tests versus conventional angiography in the same underlying population. Four studies which directly compared CACS using CT with other non-invasive tests are briefly described for context.[91, 93-95] Findings from meta-analyses of other noninvasive tests compared with CCA are provided for context.

Reliability (reproducibility) studies

From a list of 21 studies which explicitly included wording related to reliability in the title and/or abstract, three explicitly stated that symptomatic clinical patients were evaluated were identified.[86, 96, 97] These studies were moderate in quality (LoE II). In addition two LoE I/II validation study reported on reliability but did not provide adequate detail for determination of its quality as reliability study.

Safety
One study which modeled lifetime risk for radiation-induced cancer in asymptomatic persons was found and included for context.[5] One systematic review[98] and two studies with patient populations that included symptomatic persons undergoing EBCT for calcium scoring were identified both of which were contained in the systematic review.[35, 99] Two studies in asymptomatic persons referred for CACS as a screening test[36, 100]and one small study in which it was unclear whether patients were symptomatic or asymptomatic[101] were also identified. Information from these later three studies is provided for context.

Clinical decision making and patient outcomes
No studies were identified which explicitly compare a decision strategy which included CACS with an alternate decision strategy with out it were found. All were considered case series (LoE IV).

One study which describes the potential influence of CACS on referral for conventional cardiac catheterization was found.[102] Five studies which described use of CACS as a possible triage test in an emergency department setting were identified.[21, 103-106] These studies did not directly compare decisions based on CACS with a control group and are primarily case series, and thus a lower quality of evidence. It is not clear that actual decisions for discharge or further testing were actually made on the basis of the CACS.

Three studies looking at the ability of CACS to predict cardiac outcomes and mortality are briefly described.[107-109] Although the studies are of good quality and adjust for base line risk factors and other potential confounders, details of patient treatment which may influence such outcomes are not evaluated. The extent to which CACS influenced decision making which might affect these outcomes is not presented. Thus, the extent to which CACS, or any given threshold, may influence treatment and patient outcomes is difficult to assess.

Special populations
Two studies provided data comparing CACS with angiography in diabetic patients were identified and included in the listing of validation studies above. [70, 82] Both of these studies also included information on non-diabetic patients, but for one study Khaleel [70]i, there was overlap with a previously published study [77]so only the results for diabetic persons are used in this report.

Three (two LoE II, one LoE III) of the validation studies evaluated tests characteristics based on gender.[68, 110, 111] Seven LoE I/II validation studies provided information on CACS with respect to age.[66-69, 72, 74, 112]

Formal economic analyses
Two moderate quality full formal economic analyses[113, 114] were identified. One poorer quality costing study was also found and is included for context.[95]

2.4 Description of study population

Study populations in the validation studies were primarily those with symptoms of CAD. The overall prevalence of CAD determined by angiography ranged from 48.6% to 76.2%.

Table 6. Population characteristics of LoE I/II validation studies

Author (year)	Population	Inclusion criteria	Presenting symptoms
Leschka (2008)	N = 74 age: 62 (± 12) years (16-86) % male: 68	• stable clinical conditions (CCS class I-II and New York Heart Association functional class I-III)	• typical angina (n = 40) • atypical angina (n = 19) • pathological exercise test (n = 12) • dyspnea (n = 9)
Kajinami (1995)	N = 251 age: 56 (± 14) years % male: 69.3	• elective coronary angiography between May 1991 and May 1993 • chest pain on exertion or at rest or both suggesting angina pectoris • ECG findings at rest that indicated possible myocardial ischemia	NR
Nixdorff (2008)	*per-protocol* N = 71 age: 62 years % male: 59 *ITT* N = 79	• elective coronary angiography due to symptoms suspicious of CAD • primary diagnostic procedure, i.e. no previous MI, coronary intervention, or surgery	NR
Becker (2007)	N = 1347 age: 60 (± 21) years % male: 59.6	NR	• typical angina: 49% (n = 666) • atypical angina: 35% (n = 470) • exertional dyspnea: 13% (n = 175) • heart failure: 3% (n = 40)
Lau (2005)	N = 50 age: 62 (± 11) years male: 62 years (37-78); female: 61 years (36-75) % male: 80	• heart in sinus rhythm • elective conventional coronary angiography for suspected CAD	NR
Knez (2004)	N = 2115 age: 62 (± 19) years % male: 66.4	• symptomatic • referral by primary physician due to concern for possible presence of myocardial ischemia	• typical or atypical chest pain: 80% (n = 1697) • exertional dyspnea: 12% (n = 258) • heart failure: 8% (n = 160) • abnormal stress test: 52% (n = 1391)
Budoff and Diamond (2002)	N = 1851 age: 58 (± 11) years (range, 21-86) % male: 63%	• primary physician's concern for the presence of myocardial ischemia based on positive noninvasive stress testing, abnormal echocardiogram, or clinical history	NR
Lamont (2002)	N = 153 age: 58 (± 9) years % male: 76	• symptomatic patients with a positive treadmill stress test according to standard criteria who then underwent coronary angiography • all referred by primary physicians to evaluate the possibility of an ischemic cause for the symptoms	• typical angina: 37% • atypical angina: 39% • possible non-cardiac: 24%
Leber (2001)	N = 93 age: 59 (± 9) years % male: 85	• suspected CAD • chest pain with an atypical pain character, an atypical pain localization , or an unusual trigger	NR
Haberl (2001)	N = 1764 age: 20-80 years male: 56 ± 14 years female: 60 ± 16 years % male: 69	• typical or atypical chest pain and/or signs of myocardial ischemia on noninvasive tests (bicycle stress test in most cases) • clinical indication for cardiac catheterization	• "chest pain" compatible with angina: 65% • abnormal stress test: 52% (460/920)
Kwok (2000)	N = 42 age: 55 (± 10) years % male: 79	• recent MI, unstable angina pectoris, or positive stress test	• MI: 19% (n = 8) • unstable angina: 40% (n = 17) • chest pain + abnormal stress test: 40% (n =17)

NR = not reported; BMI: body mass index; CAD: coronary artery disease; CCS: Canadian Cardiac Society; DM: diabetes mellitus; HTN: hypertension; ITT: intention-to-treat, IV: intravenous; LoE: level of evidence; MI: myocardial infarction.

2.5 Description of study outcomes

The primary outcomes of interest revolve around the performance characteristics of CACS, namely sensitivity, specificity and predictive values.

- Sensitivity is the percent of persons with the disease who test positive
- Specificity is the percent of persons who do not have the disease who test negative
- The positive predictive value (PPV) is the percent of persons with a positive test who have the disease
- The negative predictive value (NPV) is the percent of persons with a negative test who do not have the disease.
- The percent of persons who have a negative test but do have disease is calculated by taking 1 – NPV.

For reliability studies, outcomes reported were degree of variation, intra-class correlation or Pearson correlation coefficient, Chronbach's alpha, and/or kappa.

Economic studies reported costs per correct diagnosis.

3. Results

3.1 Key question 1: What are the test characteristics, PPV (positive predictive value), NPV (negative predictive value), sensitivity and specificity, of coronary artery calcium scoring (CACS) compared with the reference standard of coronary angiography for the diagnosis of CAD or other established diagnostic tests for CAD. What is the evidence to describe the reliability (i.e., test-retest, intra-reader, inter-reader performance) of CACS?

Overview of validation findings

A total of 30 primary studies of accuracy and validity comparing CACS with CCA were identified. Of these, 11 studies[66-76] were classified as LoE I or II, 8 as LoE III[77-84] and 11 as LoE IV[14, 85-91]. Results from the highest quality studies (LoE I/II) formed the primary focus for analysis. Additional information on all studies may be found in the appendices.

There is not a consensus in the literature with regard to specific thresholds or cut points for what would constitute a positive versus a negative CACS test. The following thresholds were chosen for analysis: CACS > 0, ≥100 and ≥ 400. These were chosen for primary meta-analysis following consultation with a clinical expert and are based on the premise that a score of 0 might be used to "rule out" significant CAD and a score of ≥ 400 might signal the need for additional evaluation as described in the 2006 AHA scientific statement, [2] the ACCF/AHA consensus document[1] and a recent meta-analysis.[61]

Based on information from the highest quality (LoE I/II) studies, comparison of these CACS thresholds with an angiographic threshold of ≥ 50% vessel narrowing for obstructive CAD as the reference standard, these data suggests that:

- A CACS > 0 is highly sensitive (99% , CI = 98% - 99%) for detecting the presence of obstructive CAD, however 5% of persons (1 – negative predictive value) with a negative test would have CAD based on pooled estimates from seven studies with a total of N = 7354 patients. Approximately 35% (specificity) of persons without CAD might avoid unnecessary angiography or additional tests.

- Higher thresholds for CACS of ≥ 100 and ≥ 400 lowered the sensitivity (to 85% and 78% respectively) but improve the specificity (77% and 83%, respectively). Clear decisions may not be possible based on CACS when using these thresholds to define a positive test.

Although it appears that these higher quality studies followed protocols for blinded interpretation of the CACS and angiography and that the two tests were administered independent of the results of one another, these were not uniformly well stated.

The overall prevalence of CAD determined by angiography ranged from 48.6% to 76.2% across the studies used for meta-analysis. The extent to which this prevalence is characteristic of populations who would be referred for CACS as a triage test prior to angiography is not clear. While the sensitivity and specificity are generally not influenced by prevalence, the predictive values are.

Detailed results-validation studies

Patients included in the LoE I/II studies were characterized as symptomatic and/or were referred for CCA for clinical indications. An overview of patient characteristics is provided in the previous section. Not all studies provided data sufficient for meta-analysis at the cut points selected.

Meta-analysis was conducted on the primary validation parameters when data on true positive, true negative, false positive and false negative test results from three or more LoE I-II studies were available or could be calculated from author's report for the above thresholds. Not all studies reported data at these levels. All studies in these analyses used a CCA threshold of ≥ 50% reduction in luminal diameter for present of obstructive CAD. Please see the appendices for summary ROC curves.

CACS > 0

One proposed application of CACS is to use a threshold of 0 to rule out significant CAD.

Data were available from seven LoE I/II studies (a total of 7354 patients) which defined a positive test based on detection of any calcium, i.e. a threshold of > 0. The table below describes the prevalence of obstructive CAD and CACS test results based on the presence of calcium. The Lamont study included only patients who had positive treadmill stress test, which may account for the higher prevalence of CAD.

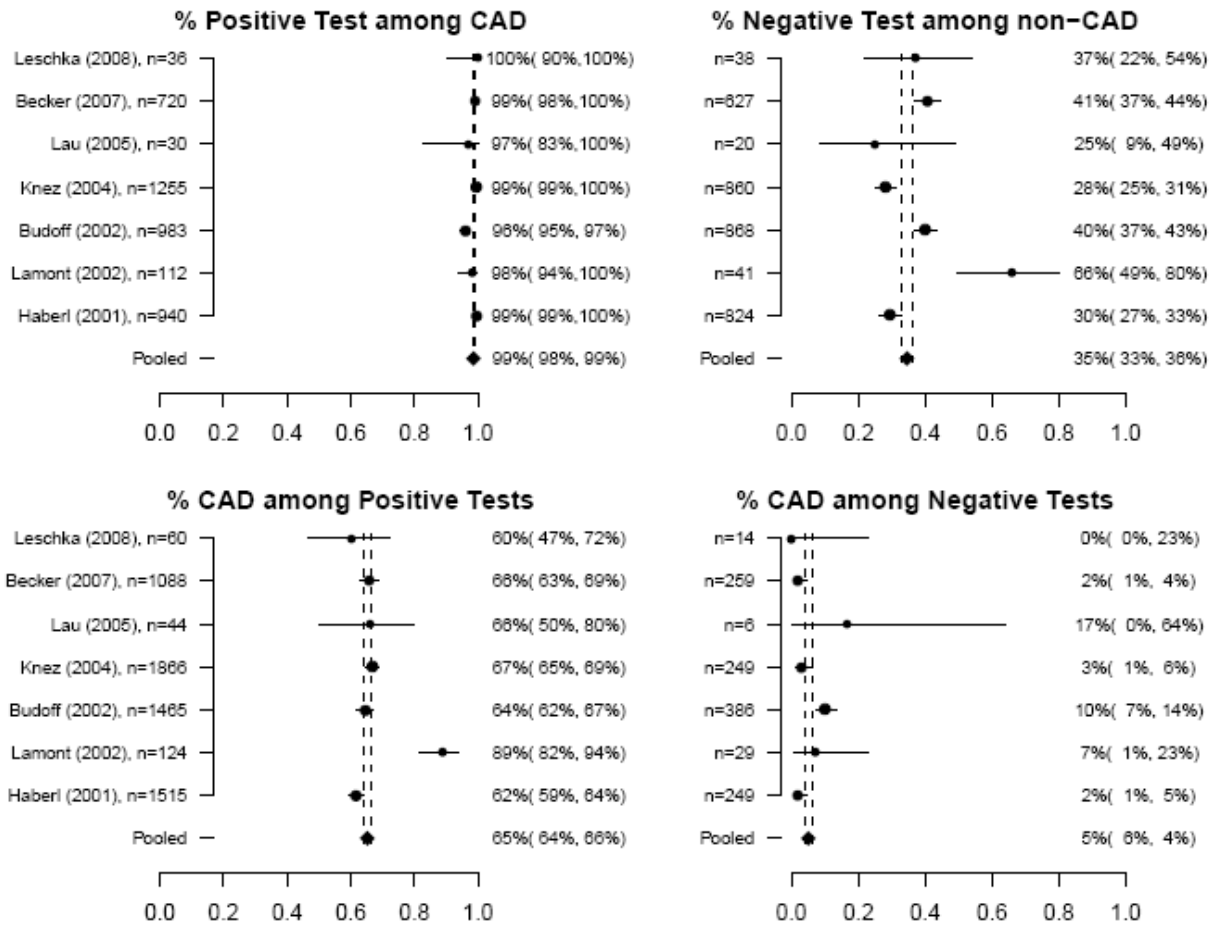
Summary test results for LoE I/II studies reporting data for CACS > 0

Author	Total N	TP, n	TN, n	FP, n	FN, n	CAD, n	CAD, %	LoE
Leschka (2008)	74	36	14	24	0	36	48.6	I
Becker (2007)	1347	715	254	373	5	720	53	II
Lau (2005)	50	29	5	15	1	30	60	II
Knez (2004)	2115	1247	241	619	8	1255	59.3	II
Budoff (2002)	1851	944	347	521	39	983	53	II
Lamont (2002)	153	110	27	14	2	112	73.2	II
Haberl (2001)	1764	935	244	580	5	940	53.3	II

CAD = coronary artery disease; FN = false negative; FP = false positive; TN = true negative; TP = true positive; LoE = Level of evidence

A summary of meta-analysis results for sensitivity, specificity, positive predictive value and 1-negative predictive value for these studies is found in Figure 3. The point estimates and 95% confidence intervals for individual studies and for the pooled estimate are given.

Figure 3. Meta-analysis of test characteristics* for LoE I/II studies reporting data for a CACS threshold of > 0 as a positive test.



*Test characteristics presented:

Upper left panel: sensitivity = % of patients with the disease who test positive
Upper right panel: specificity = % of patients who do not have disease who test negative
Lower left panel: positive predictive value (PPV) = % of patients with a positive test who have the disease
Lower right pane: 1 – negative predictive value (NPV). The negative predictive value is the % of patients with a negative test who do not have the disease so 1- NPV is the percent of patients with a negative test who do have the disease.

The sample size n refers to numbers of subjects in the denominator, that is, subjects with CAD in upper left, without CAD in upper right, with positive tests in lower left, and with negative tests in lower right.

Overall, a CACS score > 0 appears to be sensitive (99%) for detection of CAD, but the specificity is low. An estimated 5% of CAD cases would be missed based on the pooled estimate for 1-negative predictive values shown on the lower right panel of the figure.

The upper left panel shows the sensitivity of CACS > 0 as a test for the presence of obstructive CAD. The pooled sensitivity estimate is 99% with the 95% confidence interval of 98% -99%. This indicates that almost all subjects with coronary artery disease score positive on the coronary calcium test at this threshold. There is little variability in the estimates across studies.

If one use of CACS as a diagnostic test is to triage patients as candidates for conventional coronary angiography or further testing, approximately 35% of persons without CAD might avoid unnecessary angiography or additional tests based on a CACS of > 0 (specificity), however, 65% of persons without CAD would be subjected to additional testing unnecessarily (1-specificity). This pooled specificity estimate of 35% (CI = 33%, 36%) is low and variability across studies is noted.

On average, 65% of subjects (95% CI = 64%-66%) with a positive test at this threshold were found to have CAD by conventional angiography (positive predictive value) as see on the lower left panel of the above figure.

However, 5% (95% CI 4%, - 6%) of persons (1 - negative predictive value) with a negative CACS would have CAD, based on pooled estimates from the seven studies, demonstrated in the lower right panel. Missing 5% of patients with CAD may not be acceptable in a clinical setting.

The data were examined for heterogeneity across studies and statistically significant heterogeneity with standard G-squared statistics[115] for all four analyses was seen. Nevertheless, the pooled estimates seem to provide good summaries of test performance as most of the estimates from the individual studies are reasonably close to the pooled value. Large sample sizes, as were employed in several of the studies, can make small differences statistically significant. As part of the exploration of heterogeneity, analyses were repeated without the Lamont study. This study population had a higher prevalence of CAD and estimates of specificity and positive predictive value were dissimilar to estimates from the other studies. Repeating the analysis without this study had little effect on the pooled estimates: Sensitivity = 99%, specificity = 34%, positive predictive value = 65%, 1-negative predictive value = 5%.

CACS ≥ 100

Data were available from 5 LoE I/II studies representing a total of 7119 patients which defined a positive test based on a threshold of ≥ 100. The table below describes the prevalence of obstructive CAD and CACS test results based on this threshold The prevalence of CAD in the study by Kwok, et al was much higher than the others.

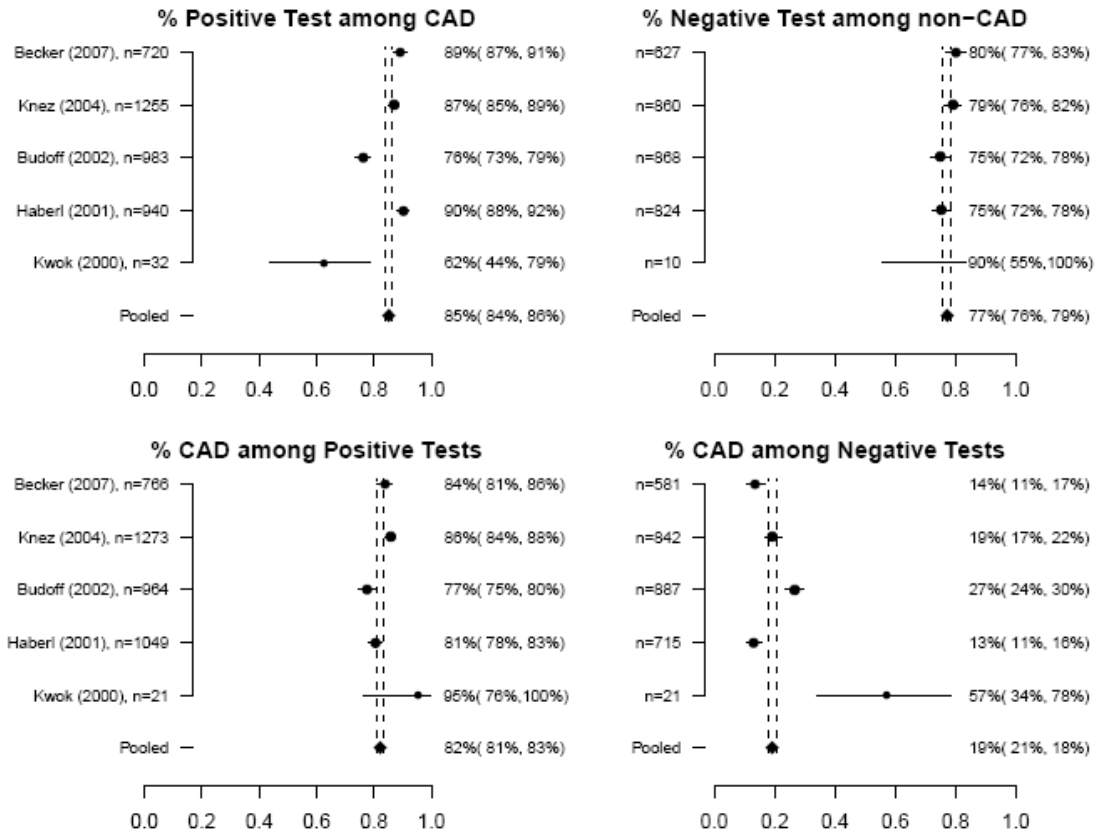
Summary of test results at CACS threshold ≥ 100.

Author	Total N	TP, n	TN, n	FP, n	FN, n	CAD, n	CAD, %	LoE
Becker (2007)	1347	641	502	125	79	720	53	II
Knez (2004)	2115	1092	679	181	163	1255	59.3	II
Budoff (2002)	1851	747	651	217	236	983	53	II
Haberl (2001)	1764	846	621	203	94	940	53.3	II
Kwok (2000)	42	20	9	1	12	32	76.2	II

CAD = coronary artery disease; FN = false negative; FP = false positive; TN = true negative; TP = true positive; LoE = Level of evidence

A summary of meta-analysis results for the same CACS test parameters using a threshold of 100 is seen in Figure 4.

Figure 4. Meta-analysis of test characteristics* for LoE I/II studies reporting data for a CACS threshold of ≥ 100 as a positive test.



*Test characteristics presented:

Upper left panel: sensitivity = % of patients with the disease who test positive
Upper right panel: specificity = % of patients who do not have disease who test negative
Lower left panel: positive predictive value (PPV) = % of patients with a positive test who have the disease
Lower right pane: 1 – negative predictive value (NPV). The negative predictive value is the % of patients with a negative test who do not have the disease so 1- NPV is the percent of patients with a negative test who do have the disease.

The sample size n refers to numbers of subjects in the denominator, that is subjects with CAD in upper left, without CAD in upper right, with positive tests in lower left, and with negative tests in lower right.

Using a CACS of ≥ 100 implies that fewer persons with CAD are detected than at the 0 cutoff (sensitivity), (85% rather than 99%), but it also increases substantially the number of persons without CAD who test negative (specificity) (77% versus 35%). In terms of decision making for individuals, a negative test at the CACS 100 cutoff does not imply that the subject is without CAD. Approximately 19% of subjects who test negative with CACS were found to have CAD (1-negative predictive value).

On average, 82% of persons with positive tests at the 100 CACS cutoff had CAD (positive predictive value). When analyses were repeated excluding the Kwok study, which had a population with higher CAD prevalence, pooled estimates were unchanged as the sample size of that study was very small relative to the others.

Again, although statistically heterogeneity was seen in these analyses, the consistency of individual study estimates with the pooled value for most studies suggests that the pooled estimates provide a reasonable summary.

CACS ≥ 400

Data were available from 3 LoE I/II studies which defined a positive test based on a CACS threshold of ≥ 400 representing a total of 195 patients. The table below describes the prevalence of obstructive CAD and CACS test results based on this threshold.

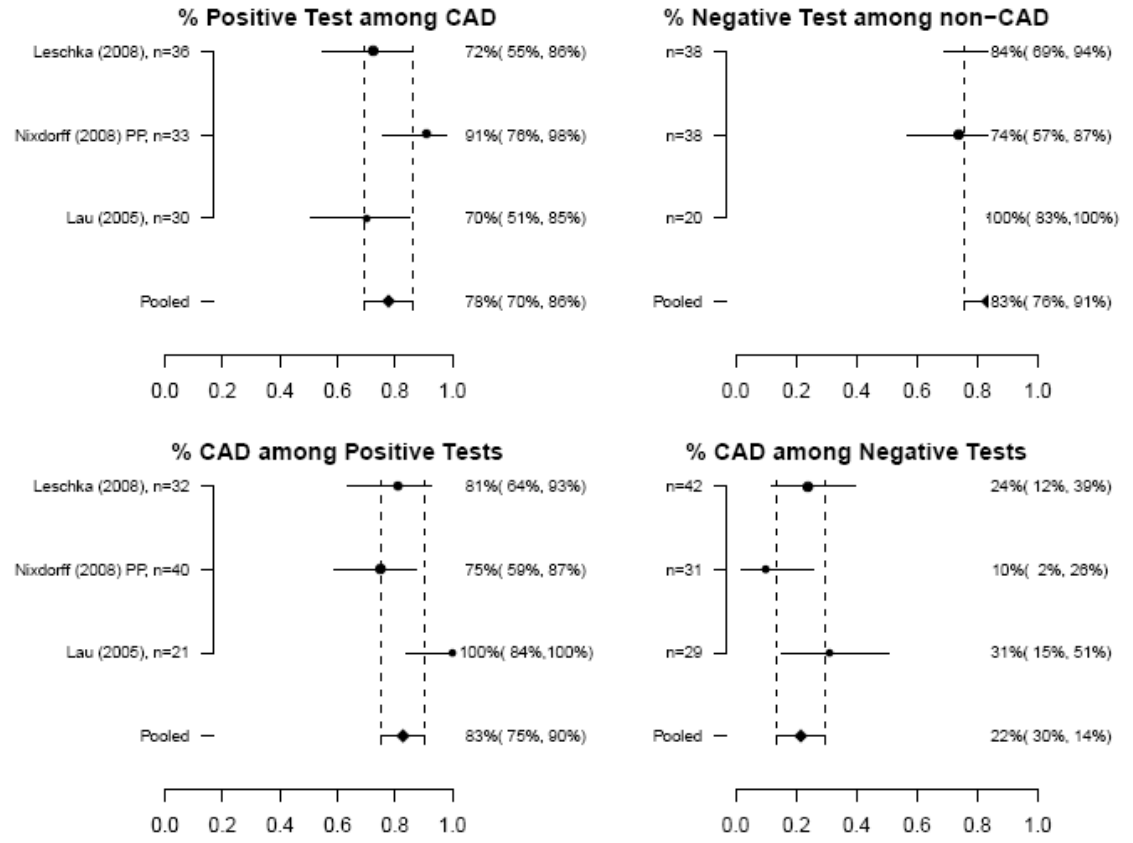
Summary of test results at CACS threshold ≥ 400.

Author	Total N	TP, n	TN, n	FP, n	FN, n	CAD, n	CAD, %	LoE
Leschka (2008)	74	26	32	6	10	36	48.6	I
Nixdorff (2008) "per-protocol"*	71	30	28	10	3	33	46	II
Lau (2005)	50	21	20	0	9	30	60	II

*the "per protocol" analysis reported appears to be based on the number of interpretable studies
CAD = coronary artery disease; FN = false negative; FP = false positive; TN = true negative; TP = true positive; LoE = Level of evidence

A summary of meta-analysis results for the same CACS test parameters using a threshold of 400 is given in the Figure 5. The point estimates and 95% confidence intervals for individual studies and for the pooled estimate are given.

Figure 5. Meta-analysis of test characteristics* for LoE I/II studies reporting data for a CACS threshold of ≥ 400 as a positive test.



*Test characteristics presented:

Upper left panel: sensitivity = % of patients with the disease who test positive
Upper right panel: specificity = % of patients who do not have disease who test negative
Lower left panel: positive predictive value (PPV) = % of patients with a positive test who have the disease
Lower right pane: 1 – negative predictive value (NPV). The negative predictive value is the % of patients with a negative test who do not have the disease so 1- NPV is the percent of patients with a negative test who do have the disease.

The sample size n refers to numbers of subjects in the denominator, that is subjects with CAD in upper left, without CAD in upper right, with positive tests in lower left, and with negative tests in lower right.

The sample sizes were comparatively small in the 3 studies that used 400 as a cutoff to define a positive test. Therefore confidence intervals for the pooled estimates of test performance are wide. As expected, the specificity estimate using the 400 CACS cutoff is even higher than at 100 (83% versus 77%), but the uncertainty is considerable and the actual specificity may be as low as 76%. Moreover, even if the estimated values of 83% for specificity and 78% for sensitivity are correct, these may not be high enough to use the test for "ruling in disease". The extent to which they may assist with referring persons on to the next level of testing is also not clear.

Looking at predictive values, between 10% and 25% of persons who test positive do not have CAD (1-positive predictive value) while 14%-30% of those testing negative have CAD (1-negative predictive value). So again, clear decisions are not possible based on CACS when using the 400 threshold value to define a positive test.

Other thresholds
A number of LoE I/II studies had data for other CACS thresholds as seen in the following table. The pattern of higher sensitivity (> 90%) is seen for thresholds up to 50.

Table 7. Summary of test characteristics at alternate thresholds

Author	Total N	CACS Cut-off	TP, n	TN, n	FP, n	FN, n	CAD, %	Sens	Spec	PPV	NPV
Becker (2007)	1347	>10	698	408	219	22	53.0	97%	65%	76%	95%
Knez (2004)*	2115	> 10	1180	602	258	75	59.3	94%	70%	82%	89%
Budoff (2002)	1851	> 20	885	503	365	98	53.0	90%	58%	71%	84%
Haberl (2001)	1764	≥ 20	914	531	293	26	53.3	97%	64%	76%	95%
Leber (2001)	93	> 46	40	27	22	4	47.3	91%	55%	65%	87%
Lau (2005)	50	≥ 50	27	11	9	3	60.0	90%	55%	75%	79%
Budoff (2002)	1851	> 80	777	625	243	206	53.0	79%	72%	76%	75%
Leber (2001)	93	≥ 130	35	35	14	9	47.3	80%	71%	71%	80%
Kwok (2000)	42	≥ 160	16	9	1	16	76.2	50%	90%	94%	36%
Leber (2001)	93	> 310	25	43	6	19	47.3	57%	88%	81%	69%

*based on volumetric score; CAD = coronary artery disease; FN = false negative; FP = false positive; TN = true negative; TP = true positive; Sens = sensitivity; Spec = specificity; PPV = positive predictive value; NPV = negative predictive value

LoE III validations studies
Seven LoE III studies had sufficient data for meta-analysis comparing a CACS threshold of > 0 with the referent standard of angiography (≥ 50% vessel narrowing).[77-79, 82-84]. Additional information is found in the appendices.

In five of the 6 LoE III studies used in the meta-analysis, blinded interpretation of the CACS and angiograms was reported. In these studies, however, it was not clear that the decision to perform the test and the referent was independent. In one study only 71% of persons who had CACS also had angiography.[78] These factors increase the potential for bias. One study which used an angiographic threshold of ≥ 75% was not included in the meta-analysis.[81] Additional meta-analysis information for LoE III studies is in Appendix E.

Table 8. The LoE III studies for which data at CACS > 0 were available are listed below.

Author	Total N	TP, n	TN, n	FP, n	FN, n	CAD, n	% CAD
Hosoi (2002)	181	95	26	41	6	114	63
Chen (2001)	116	63	23	29	1	64	55.2
Shavelle (2000)	97	66	14	16	3	69	71
Budoff (1996)	710	404	124	159	23	427	60
Tannenbaum (1989)	54	38	11	0	5	43	80
Herzog (2004)*	38	17	4	16	1	18	47
Fallavollita (1994)	212	100	42	52	18	118	56%

*Used angiographic cut off of ≥ 75% and was not included in meta-analysis
CAD = coronary artery disease; FN = false negative; FP = false positive; TN = true negative; TP = true positive;

Pooled estimates and 95% confidence intervals for test parameters from the LoE III studies are compared with those from LoE I/II studies as follows:

	LoE III studies	LoE I or II studies
Sensitivity	93% (92, 95%)	99 % (98%, 99%)
Specificity	44% (40, 48%)	35% (33%, 36%)
Positive predictive value	71% (69%, 74%)	65% (63%, 66%)
1 – negative predictive value	19% (15%, 23%)	5% (4%, 6%)

The pooled sensitivity estimate for the lower quality studies is somewhat lower compared with the LoE I/II studies and the specificities somewhat higher. The prevalence of angiographically detected CAD in the LoE III studies ranged from 47% to 88% and most had higher prevalence than did the LoE I/II studies. This may partially explain the differences in predictive values.

Only one LoE III study with 181 persons without diabetes provided data using thresholds of 100 and 400 and compared CACS with an angiographic cut off of ≥ 50% luminal narrowing. [Hosoi] The prevalence of CAD was 63%. Test characteristic estimates (95% CI) for CACS of 100 and 400 respectively were:

- Sensitivity: 66% (56%, 74%) and 43% (34%, 53%)
- Specificity: 83% (74%, 92%) and 97% (92%, 100%)
- PPV: 87.2% (80.2%, 94.3%) and 96.2% (90.9%, 100%)
- 1 – negative predictive value: 41.1% (31.2%, 50.9) and 49.6% (41.0%, 58.2)

General findings from LoE IV studies can be found in Appendix F.

Comparison with other diagnostic tests

The validation studies included in this technology assessment used coronary angiography as the reference standard. Since CACS is noninvasive test, it may be helpful to compare its performance characteristics with other non-invasive tests. Since these are not the appropriate reference standard, test characteristics are not reported for comparisons of CACS with them. These other noninvasive tests reflect cardiac physiology and function. [77] Such noninvasive tests include those that use exercise or drugs (such as dobutamine) to stress the heart to evaluate myocardial function and the extent to which it may have been compromised by obstructed coronary artery blood flow. By contrast, CACS provides anatomic information about calcium amount and distribution, thus, the basis for CACS is different than that for other noninvasive tests. Aspects of cardiac function may be assessed by ETT, echocardiography, or imaging radionuclide tracers. The information in this section is presented to provide context regarding these tests.

Test performance characteristics reported in meta-analyses comparing exercise stress testing (ETT, echocardiography or nuclear perfusion) with angiography are briefly summarized below. For comparison, pooled estimates from the meta-analysis presented for LoE I and II studies in this technology assessment are also listed.

Study	Noninvasive test	Sensitivity	Specificity
Gianrossi [11] (N =24,074 patients; 141 studies)	Exercise ECG	68% (range 23% - 100%)	77% (range 17% - 100%)
Fleischmann[116] (N = 2637; 24 articles*)	Exercise Echo	85% (95% CI 83%, 87%)	77% (95% CI 74%, 80%)
Fleischmann[116] (N = 2637patients; 27 articles*)	Exercise SPECT	87% 95% CI 86, 88)	64% (95% CI 60%, 68%)
Present Spectrum Research HTA (N = 7354 patients; 7 studies)	CT CACS score > 0	99% (95% CI 98%, 99%)	35% (95% CI 33%, 36%)
(N = 7119 patients; 5 studies)	score ≥ 100	85% (95% CI 84%, 86%)	77% (95% CI 76%, 78%)
(N = 195 patients ; 3 studies)	score ≥ 400	78% (95% CI 86%, 70%)	83% (95% CI 76%, 91%)

* not unique patient data sets

Two validation studies (one LoE I and the other LoE III) included in this HTA also compared other non-invasive tests to angiography in the same patient population as that used for assessing CACS.[69, 83] As opposed to studies comparing an individual non-invasive test to angiography, different patient characteristics cannot explain differences between test performance measures. A summary of test performance characteristics for CACS and other noninvasive tests compared with angiography in the same patient population is provided below.

	Sensitivity %	Specificity %	PPV %	NPV %
Kajinami 1995 (LoE I)				
Thallium exercise	83	60	70	76
Stress echo	74	73	77	72
CT calcium score > 19	77	86	86	76
Shavelle 2000 (LoE III)				
Treadmill-ECG	76	60	81	53
Technetium exercise	78	67	83	57
CT calcium score > 0	96	47	80	82
Calcium score > 0 and abnormal treadmill-ECG	72	83	91	57

NPV is negative predictive value; PPV is positive predictive value

Four studies which compared CACS directly with other noninvasive tests were identified and varied in quality.[91, 93-95] Findings from these other tests at various CACS thresholds are presented below:

Study	Population	Comparison	Ca score cut-off	Findings from comparison test

Janssen 2005	114 outpatients with chest pain and inconclusive clinical findings, rest ECG, and exercise ECG test	Dobutamine magnetic resonance imaging	< 11
			
			and
			<100 vs
			≥ 100
				n = 29 patients with CACS <11
				• none of the 29 had an abnormal stress MRI
				n = 46 at CACS < 100
				• # with abnormal stress MRI test
				n = 68 at CACS ≥ 100
				• 20 with abnormal stress MRI test

Nishida 2005	83 patients suspected of having ischemic heart disease	Thallium exercise stress test	0 vs. > 0	n = 42 patients with CACS = 0 • 12 had an abnormal thallium scan n = 41 patients with CACS > 0 • 23 had an abnormal thallium scan
Raggi 2000	207 patients with chest discomfort	Exercise treadmill test	0 vs. >0 and <150 vs ≥ 150	n = 82 patients with CACS = 0 • 7 had a positive treadmill test • 23 had an equivocal treadmill test n = 172 patients with CACS <150 • 16 had a positive treadmill test • 52 had an equivocal treadmill test n = 25 patients with CACS ≥ 150 • 4 had a positive treadmill test • 11 had an equivocal treadmill test
Yao 2004	73 clinically stable patients suspected of having coronary artery disease, no history of MI; only 30 had chest pain	Technetium SPECT	0 vs. >0	n = 29 patients with CACS = 0 • 7 had an abnormal SPECT n = 44 patients with CACS > 0 • 24 had an abnormal SPECT

Reproducibility, reliability

Studies in symptomatic patients that evaluated reliability for total calcium scores for patients (versus by vessel or segment) were sought. Since the scoring method described by Agatston in 1990 is the most widely used and reported in the validation literature, only those studies that evaluated the reproducibility of this method were included. Studies comparing slice thickness, contiguous versus overlapping methods, different software or assessment protocols were excluded.

Use of symptomatic subjects was apparent in only three studies of reliability.[86, 96, 97] In an examination of eleven LoE I and II validation studies, two studies provided limited information on reliability.[73, 75] The overall quality of reliability studies of CACS in symptomatic patients was moderately high; all were LoE II.

Assessment of level of evidence (LoE) for reliability studies on coronary artery calcium scoring (CACS)

Methodological Principle	Broderick (1996)	Möhlenkamp (2001)	Serafin (2009)
Broad spectrum of patients with expected condition		■	
Adequate description of methods for replication	■	■	■
Blinded comparison of tests/interpretations	■		■
Evidence Level	**II**	**II**	**II**

Evidence from reliability studies in symptomatic subjects suggests that calcium scoring has moderate to good reproducibility in test-retest studies and high interobserver agreement in an inter-rater study.[86, 96, 97] Two LoE I or II validation studies reported moderate to excellent interobserver agreement between raters but did not provide detail of study design regarding reliability evaluation. [73, 75]

Methods for determining Agatston score reliability were:

- Intraclass correlation coefficient: assesses the consistency of measures made by multiple observers measuring the same quantity. Higher values indicate good consistency of measures, with a maximum value of 1.0.
- Variability of Agatston score: indicates the relative difference between scores of two tests in a test-retest study [Bland]. Lower values indicate less variability (more reproducibility) in the scoring of the two tests. The calculation is:

$$\frac{|\text{score from scan 1} - \text{score from scan 2}|}{\text{Average score from both scans}} * 100$$

Table 9. Overview of primary findings for reliability studies on coronary artery calcium scoring (CACS).

Author (year)	Type of study	Calcium score	Measure of reliability	LoE
Broderick (1996) N = 101	Test-retest Inter-rater	NR	Intraclass correlation coefficient = 0.90 (test-retest), 0.99 (inter-rater)	II
Möhlenkamp (2001) N = 50	Test-retest	NR	Variability of Agatston score = 21.8% (mean), 19.2% (median)	II
Serafin (2009) N = 50	Test-retest	Median = 511.3	Variability of Agatston score = 3.9% (median)*	II
Leschka (2008)	Inter-rater	Mean = 720 ± 968 (0 – 4387)	kappa = 0.84	*
Lau (2005)	Inter-rater	NR	Intraclass correlation coefficient = 1.00	*

* Leschka and Lau are LoE I/II validation studies which also reported reliability information. Since they did not describe how reliability was evaluated, it is not possible to provide a LoE determination for this.

In the three studies in which it was clear that symptomatic persons were enrolled, the test-retest reliability was moderate. Various methods were used to evaluate this.

- The reproducibility of the total calcium score as measured by the intraclass correlation coefficient was good (0.90), however this measurement applied only to 17 subjects[86] based on method and algorithm that mimics the Agatston method for calcium scoring.
- Another study reported that 86% with a CACS = 0 in one scan also had a a score of 0 zero in the second. Moderate variability in the calcium score from test to retest (21.8%, 19.2% for the mean and median calcium score, respectively), which decreased with increasing amounts of calcium was reported in another study. [96] However, the area score exhibited lower variability than the calcium score and also decreased with increasing amounts of calcium. The definition of "moderate variability" was not provided.

Evidence from three studies suggests moderate to high inter-observer agreement between raters of calcium scores.

- Interobserver agreement of two raters as measured by the intraclass correlation coefficient was high (0.99).[86]
- Moderate interobserver agreement (0.84) as measured by the kappa statistic was found in a study that focused on the validation of calcium scoring and computed tomography coronary angiography compared to convention coronary angiography.[75]
- Excellent interobserver agreement (1.00) as measured by the intraclass correlation coefficient was found in a study that focused on the validation of calcium scoring and CT angiography compared with conventional angiography.[73]

Factors that might influence reproducibility of calcium scoring include image noise, the number of images acquired, lesion size, overall extent of calcium, motion artifacts and ECG gating, mistriggering, arrhythmia, table movement, patient movement, breathing, and heart rate.

3.2 Key question 2: What is the evidence related to the safety of CACS

The two primary safety issues related to CACS are radiation exposure and the observation and evaluation of incidental findings. Unlike coronary computed tomography angiography, CT calcium scoring does not require a contrast agent, and so avoids the risks of contrast reactions and drug-induced nephropathy.

Radiation exposure
The main safety concern regarding calcium scoring by CT is that it exposes the patient to low to moderate levels of ionizing radiation, which may increase the life-time risk of cancer. To the extent that CACS reduces the need for conventional angiography, exposure might be reduced. To the extent that CACS results in the need for additional testing, it may be increased. To date, no large-scale epidemiologic studies evaluating cancer risk associated with CT in general have been published. A recently published study estimated risks related to CACS based on radiation risk models for cancer incidence.[5] There is uncertainty and controversy with regard to the actual risk of low dose radiation. Quantification of risk specific to CACS for an individual patient is not possible.

The American College of Radiology's (ACR) 2008 Appropriateness Criteria on the evaluation of chronic chest pain in patients with a low to intermediate probability of CAD lists the relative radiation level of CACS as medium, between 1-10 mSv. In clinical decision making, this level of exposure needs to be put in the context of other tests which also may involve radiation that may be part of the clinical pathway, as the possible cumulative effects of multiple procedures are of concern.

The term radiation exposure may refer to the measureable quantity of ionizing events in air that are produced by x-rays. By contrast, radiation dose describes the quantity of radiation energy deposited in a person as a result of the exposure. Radiation dose may be expressed as the "equivalent effective dose" in units of Sieverts (SV) or milli-Seiverts (mSv). It is typically calculated or modeled based on estimates of energy absorption per body mass unit (e.g. kilograms of body weight) and other factors. While it does not represent the dose received by an individual patient, it does provide a common metric by which different sources of radiation might be compared. [31]

To provide some context, estimates of typical effective dose for environmental and medical sources of radiation are outlined below. Some radiation exposure occurs naturally and during activities of daily living. As seen below, estimated dose for CACS ranges from 0.7 mSv to 12 mSV, based on information from various sources. Estimates include EBCT and MDCT and data may combine prospectively triggered and retrospective gating.

Table 10. Overview of typical effective dose for various radiation sources

Exposure type	Typical effective dose (millisieverts)
Environmental Exposures	
Natural source (average US per year)	3
Round trip cross-country air flight	0.02-0.05
Nuclear power plant worker	3
Exposures from diagnostic radiology	
Dental X-ray	0.005
Chest Xray (PA and lateral)	0.1
Cervical spine X-ray	0.2
Mammogram	0.4
Lumbar spine X-ray	1.5
Head CT	2
CT calcium scoring	3
Range found in validation studies in this report*	(1.2 —10)
Range found in literature 1980-2007 [Mettler]	(1—12)
Range reported in 2006 AHA Statement [Budoff]	(0.7 – 1.9)
Interventional coronary angiography	7
Barium enema with fluoroscopy	8
Virtual colonoscopy	10
Chest CT for pulmonary embolism	15
CT coronary angiography	16

*A list of studies and reported exposures is found in the appendices
Compiled from Mettler 2008; FDA—What are the radiation risks; FDA—Quantities and units; NRC; DOE; Budoff 2006[2, 30, 117, 118]

The extent to which levels of radiation used for medical procedures increase cancer is unclear. Most data are from the atomic bombing of Japan and nuclear accidents, with risk estimates for low dose exposure extrapolated back from risks for those high dose exposures. Extrapolation of these results to lower doses used in medical imaging is the subject of much controversy. However, data are also available from Japanese who received low doses from the atomic bombs and from people who receive low doses during occupational exposure. [5, 117, 119]

Two different hypotheses have been advanced for evaluating potential risk for cancer at low radiation doses. The linear quadratic approach states that malignancy risk is so low at low radiation does that it is nearly impossible to quantify but that there is quadratic increase in risk with increasing dose. The *linear no-threshold hypothesis* implies that extrapolation of malignancy risk at high doses is reasonable to situations with low doses. It states that there is no threshold below which radiation cannot cause malignancy and that the risk increases linearly with increasing dose.[33] This latter approach is more conservative and is the one generally followed. Current guidance from regulatory bodies is that no threshold exits and that exposure should be kept as low as reasonably achievable (ALARA).

In the US, 1 in 5 people will die from cancer. A CT exam delivering a radiation dose of 10 mSv may increase that risk by 1 in 2000, or by 0.05%, based on estimates extrapolated from A-bomb survivors. For comparison, approximately 400 out of 2,000 individuals are expected to develop cancer from all other sources combined. [117] While the increased risk for an individual may be considered low, the potential increased risk for the entire population creates many more cases of fatal cancer, especially as the use of CT scans expands. At the rate of CT use from 1991-1996, about 0.4% of all cancers in the US might have been due to radiation exposure during CT scans. By 2006, 1.5 to 2.0% of cancers in the US might have been due to radiation exposure during CT scans. [34]Also, as CT scans are being used for younger patients and as life expectancy increases, there are longer life spans during which CT scans might induce cancer. [34]

The radiation exposure reported during calcium scoring varies more than 10-fold. [5, 29, 30] Several factors influence the radiation exposure; the model of the CT scanner, the scan mode (e.g., axial or spiral), the voltage and current used, the number of scans, the scan pitch (the overlap between CT slices), the slice thickness, ECG triggering or gating, scanning time, the length of the patient's body that is scanned, and the patient's size. Multi-detector CT gives more radiation exposure than electron beam CT. Higher radiation doses and longer times result in higher exposure. With ECG gating, the CT beam is on taking an image only at a specific point in each cardiac cycle to avoid variation caused by cardiac motion. However, with retrospective gating, the CT beam is on, taking images throughout the cardiac cycle; images at the same points in the cardiac cycles are recreated later. While it allows the person interpreting the scans greater flexibility, retrospective gating gives more radiation exposure than prospective gating. Smaller patients received higher doses because they have less tissue to attenuate radiation. Some factors influence the quality of the images obtained, and the need for

good-quality images must be balanced against the need to minimize radiation exposure.[2, 5, 34]

To reduce radiation exposure during calcium scoring by CT, the American Heart Association recommends using prospective ECG gating, slice widths of 2.5 to 3 mm, and radiation doses adjusted for body size.[2] As the technology and techniques improve, radiation exposure could decrease. In reviewing literature published between 1980 and 2007, Mettler et al found a typical radiation exposure of 3 mSv with a range of 1.0 to 12 mSv during calcium scoring.[29] However, using equipment, software, and protocols that were up-to-date in 2003, Hunold reported radiation exposures during calcium scoring using electron beam CT of 1.0 mSv for men and 1.3 for women; and using different multi-detector CTs, 1.5 – 5.2 mSV for men and 1.8– 6.2 mSv for women based on measurements from an anthropomorphic phantom.[120]A recent AHA scientific advisory suggests that with prospective gating, the effective dose is estimated at 1 mSV and gives a range of 0.7 – 1.9 mSV.[2]

A recent simulation estimating radiation dose and cancer risk suggests that a single scan for CACS may increase lifetime cancer risk.[5] The authors used a Monte-Carlo simulation based on protocols from three clinical settings and risk models derived from Japanese atomic bomb survivors and medically exposed cohorts. While their model was for screening of asymptomatic persons, it may be reasonable to consider their results for a single CACS determination as applicable to a diagnostic scan in symptomatic patients. A single screening test at age 40 would increase the risk of cancer by 9 per 100,000 for men and 28 per 100,000 for women. For a single screen at 55 years of age, based on a median effective dose of 2.3 mSv, site specific estimates for life-time risk of radiation induced cancer suggest that most cases would be lung cancer (6/100,000 in men, 14/100,000 in women) or breast cancer (4/100,000 in women). Risks are higher for women than for men because of the radiation dose to breast tissue.

The ACR document rates the relative radiation level for CT angiography of the coronary arteries as high. The ACR range for high relative radiation level is 10-100 mSv and CT angiography is at the lower end of that range (2.0 mSV – 16.0 mSv). However, to the extent that CACS would become a routine part of CT angiography, there is the potential for greater radiation exposure in part due to the shift from EBCT to multi-detector CT.

There is potential for increased risk secondary to radiation exposure in general based studies of the atomic bombing of Japan, nuclear accidents and occupational exposures. Quantifying the explicit risk for a specific test like CACS either alone or combination with other tests involving radiation is not possible. While simulation and modeling of the effects of radiation exposure provide important insights into the possible changes in risks, verification with epidemiologic studies presents challenges since many factors which may influence development of malignancy need to be considered such as time for development, presence of additional risk factors and other exposures. The true attributable risk from radiation-based diagnostic tests may be difficult to determine. Some experts consider the potential for harm from radiation exposure to be clinically significant particularly given that patients may be likely to have additional tests using radiation. Decision making between physician and patient should involve a discussion of

the possible risks and potential benefits of CACS (and subsequent testing). Final determination of net benefit for a given clinical scenario reflects the values and judgments of the persons making the decisions.

Incidental findings

Abnormalities may be identified that are unrelated to the reason for getting the imaging study. During the CT scan for calcium scoring, parts of the lung, aorta, chest wall, breasts, spine, skin, and upper abdomen are exposed to radiation and imaged along with the heart. [35, 99] A small field of view, focusing on the heart, improves resolution for interpreting the cardiac images. But the field of view may be expanded to reveal these other areas for examination.[101] The number of incidental findings identified is affected by the scanner used, the slice thickness, and the area imaged. Identification of incidental findings may have benefits as well as drawbacks. Data from two studies of symptomatic persons, suggests that 7%-10% of them will have incidental findings during a CT scan for calcium scoring that require further diagnostic testing and a small percent, 1.2%, will require therapeutic intervention.

One systematic review[98] and two studies with patient populations that included symptomatic persons undergoing EBCT for calcium scoring were identified both of which were contained in the systematic review.[35, 99] Two studies in asymptomatic persons referred for CACS as a screening test[36, 100]and one small study in which it was unclear whether patients were symptomatic or asymptomatic[101] were also identified. Information from these later three studies is provided for context.

The quality of reporting in these studies was variable. Slightly different criteria for which incidental findings were considered clinically important were used across studies. Varied methods for identifying such findings were used and included counting of incidental findings from exam reports and re-reading of scans to identify incidental findings. The method used may influence the prevalence of such findings.

Not all incidental findings are clinically important. Clinically important findings affect patient management. Hunold reported the prevalence of incidental finding with therapeutic consequences as well as those with only diagnostic implications. However, in most studies, clinically important incidental findings are defined as those needing further testing or follow-up.[98] Table 11 provides and overview of study findings. Additional information is found in Appendix G.

Table 11. Summary of studies reporting incidental findings on CT for calcium scoring

Reference	CT characteristics	Population	n	Subjects with clinically important incidental findings n (%)
Hunold 2001	EBCT; 3 mm slices; pulmonary arteries to apex; 32% also had contrast and CT angiography*	9.9% screening exams, others had known or suspected coronary artery disease; age 20-86 y; smoking status not reported	1812	191 (10.5%) had findings with diagnostic implications; 22 (1.2%) had findings with therapeutic consequences
Horton 2002	EBCT; 3 mm slices; pulmonary arteries through apex	Symptomatic and asymptomatic ; age 23-87 y; 25% current or former smokers	1326	103 (7.8) had findings requiring clinical or radiological follow up
Elgin 2002	EBCT; 3 mm slices; axial length not described	Asymptomatic; age 39-46 y; 13% current smokers	1000	54 (5.4%) had findings requiring follow up or additional testing
Schragin 2004	EBCT; 3 mm slices; aortic root through apex	Population screening; age 21-86 y; 41% current or former smokers	1356	57 (4.2%) had findings requiring further evaluation
Law 2008	MDCT; 3 mm slices; carina to apex	Symptomatology not described; mean age 56 y; 17% current or former smokes	140	11 (8%) had findings requiring clinical or radiological follow up

*Authors do not separate out incidental findings for CACS and CT angiography
EBCT is electron beam computed tomography; MDCT is multidetector computed tomography; y is years

Many of the risk factors for coronary artery disease (such as increased age, smoking, and male sex) are also risk factors for other diseases (such as lung cancer) that might be identified on CT scans.[98] The prevalence and type of incidental findings from CT scans may vary according to the study cohort's risk factors and whether symptomatic or asymptomatic persons were included. These differences may partly explain the variation in prevalence of incidental findings between studies.

In the two studies in symptomatic persons, the majority of extra-cardiac findings requiring further testing were pulmonary nodules that were considered suspicious for tumor. [35, 99] Only Hunold reported the number of findings with therapeutic consequences.[99] Eleven extra-cardiac findings (0.6% of all findings, 9.4% of the 1175 thoracic or abdominal findings) required therapeutic intervention. The same authors also reported on 136 cardiac related findings (7.5% of all incidental findings, 20% of the 676 cardiac findings) required further evaluation, 11 of which required therapeutic intervention. Calcification of the mitral or aortic valve was the most common finding requiring additional testing and masses in a cardiac chamber were the most common findings requiring therapeutic intervention. Horton did not report on findings requiring therapeutic intervention.

The reporting of an incidental finding often leads to additional diagnostic testing. For example, radiological studies with better spatial resolution such as spiral CT with IV contrast may be done.[35, 99] These may increase overall patient radiation exposure. The additional diagnostic tests raise patient anxiety, have risks, and cost money. The additional testing may identify disease at an early stage and lead to necessary treatment, but may not actually improve survival or other patient outcomes. An apparent improvement in survival may be simply due to lead-time bias. Also, especially among low-risk patients, the incidental finding may not require treatment: the additional testing may be done with no change in the patient's outcome.

3.3 Key Question 3: What is the evidence that CACS influences clinical decision making and improves patient clinical outcomes (e.g. mortality)

Clinical decision making

The role of CACS as a diagnostic test in symptomatic patients is unclear making its influence on clinical decision making difficult to assess. Some authors have suggested that CACS may facilitate triage of patients separating out those in whom the likelihood of significant CAD is low from those who are more likely to have significant CAD and should be evaluated with additional testing.

One study from a tertiary referral center [102] and five studies describing the use of CACS as a tool for patient evaluation and decision making in the emergency department (ED) were identified. [21, 103-106] There was wide variation in the definitions of outcomes and use of CACS. None of the studies identified used concurrent or historical controls, making it difficult to effectively evaluate the impact of clinical decisions made based on CACS on patient outcomes compared with other decisions. If CACS was a perfectly sensitive test, there were no false negatives and some degree of specificity, the benefit of doing CACS as a first test for triage could be estimated in the absence of an explicit comparative group. However, without these features or a comparison group, the benefit of CACS as a first test is not clear. The studies are briefly described below. Additional detail is found in Appendix G.

Piers [102] retrospectively evaluated 598 patients with no prior CAD history with a mean estimated 10 years risk of CAD of 6% (8%-10%) who had CACS were divided in four groups based on CACS score. Decisions to do additional tests for evaluation of myocardial ischemia were made by the treating cardiologist and if judged clinically relevant, conventional angiography was performed. Overall 89% (n = 531) underwent tests for ischemia evaluation and 24% had conventional angiography with the majority of these having had tests for ischemia prior to angiography. The authors reported that the probability of referral to CCA increased with increasing CACS (P< 0.001) and that CACS may facilitate decision making.

	Ischemia test done*	**CCA done**	**CAD on CCA†**
CACS < 10 (n =304)	263 (86%)	1 (0.3% of 304)	1 (0.03%)
CACS 10-99 (n = 135)	122 (90%)	30 (22%)	13 (9.6%)
CACS 100-399 (n = 89)	85 (96%)	30 (34%)	10 (11%)
CACS ≥ 400 (n = 70)	61 (87%)	60 (86%)	35 (50%)

*Exercise stress test, ST segment analyses on 24 ECG registration or SPECT at clinician discretion;
†not all patients had CCA the number of persons with CAD who did not have CCA is unknown.

From this study it is not clear to what extent CACS as a stand alone test truly affected the decision to refer patients for angiography since 89% of patients had other non-invasive tests prior to catheterization which may have influenced the decision to perform angiography.

Studies from emergency department settings

Five studies were performed in patient populations admitted to emergency departments (ED) with angina-like chest pain and normal or nondiagnostic ECGs and/or normal cardiac enzymes.[21, 103-106] The spectrum of patients described in these studies was primarily described as low to intermediate risk. Patients were excluded if they had a prior history of coronary artery disease (CAD). Since these studies are not validation studies and are considered only for the purpose of evaluating the influence of CAC scores on clinical decision-making practices, sensitivity and specificity are not reported.

These studies vary in quality and as previously stated, did not employ a comparison group and are considered case series. These studies explore the potential for CACS as a triage tool by looking at the extent to which various outcomes were associated with CACS. Only one study, Geluk, specified an algorithm for decision making based on CACS. The decision to discourage follow-up appears to have been based on combination of exercise testing and CACS, which were performed in random order. In some cases tests were performed after discharge. It is not clear to what extent the actual decision for discharge was based algorithm described.

Authors suggest that the absence of calcium or very low score, usually < 10, may serve as an important diagnostic threshold, allowing physicians to confidently discharge such patients to home without further work-up, serving as a type of triage method for this specific patient population. Based on these studies, clinicians may use high calcium

scores as one factor in deciding which patients need intervention (such as diet, exercise, antiplatelet therapy, or cholesterol-lowering agents) or further cardiac testing. This has not been explicitly evaluated in any of the studies identified.

Geluk [104] prospectively evaluated low-risk patients with symptoms of CAD, normal ECG, and normal troponin to determine the efficacy of CAC scores compared with exercise testing. All patients underwent both calcium scoring by EBCT and exercise testing (or myocardial perfusion scintigraphy for the 7% unable to exercise). The protocol for further treatment depended on the calcium score. Patients with a CAC score of < 10 would be discharged to home and follow-up visits discouraged. Those with scores 10-399 would begin primary prevention measures such as life style modification and pharmacotherapy and additional testing at the judgment of the treating cardiologist. Patients with a CAC score ≥ 400 would have coronary angiography. Patients were followed-up after at least 4 months by review of medical records, phone interviews, or phone contact with the patient's general practitioner. The endpoint was a combination of ≥ 50% stenosis on angiography, revascularization, myocardial infarction, or cardiac death during follow up. Although only 27% of patients had angiography, all patients were followed up. Among those 159 patients with a calcium score < 10, the stress test was negative in 113, positive in 15, and nondiagnostic in 31. Despite the protocol, 13 had angiography, which did not show any obstructive lesions. None of the 159 patients with a calcium score < 10 had the combined endpoint. Among those 103 patients with a calcium score 10-399, the stress test was negative in 63, positive in 9, and nondiagnostic in 31. Thirty-three of the 103 had angiography, which showed obstructive lesions in 14 (14%), requiring revascularization in 9 (9%). All patients with coronary artery disease in this group received pharmacologic therapy. Among those 103 patients with a calcium score 10-399, 14 (14%) had a combined endpoint. Among those 42 patients with calcium score ≥ 400, the stress test was negative in 15, positive in 13, and nondiagnostic in 14. All of the patients with a calcium score ≥ 400 had pharmacologic therapy, which resolved symptoms for five. The other 37 had angiography, which showed obstructive lesions in 24 (57%) requiring revascularization in 17 (40%). Among those 42 with calcium score ≥ 400, 24 (57%) had a combined endpoint. The authors suggest that the calcium score may be used as a "gatekeeper" for additional invasive and noninvasive testing, providing effective triage in patients with suspected but low risk of CAD. Furthermore, the authors indicate that CACS is diagnostically superior to exercise testing and is a better predictor of future cardiac events.

Georgiou [105] investigated the association between EBCT detected CAC and future cardiac events in a prospective observational study of 192 patients admitted to the ED of a large tertiary care hospital for chest pain with a normal or nondiagnostic ECG. Treating physicians and patients were not told the calcium score. Outcomes were "hard events" (cardiac death or nonfatal MI) or "total events" (cardiac death, nonfatal MI, coronary revascularization, ischemic stroke, or hospitalization for angina) as ascertained by review of hospital records at a mean of 50 months after admission. The presence of calcium (CAC score > 0) was strongly related ($P < .001$) to the occurrence of hard cardiac events

(death, myocardial infarction) and all cardiovascular events (death, myocardial infarction, coronary revascularization, ischemic stroke, subsequent hospitalization for angina). CAC scores ranged from 0 to 4607 and results were categorized by quartiles of calcium scores with 48 patients in each quartile: 0, 1 to 4, 5 to 332, and 333 to 4607. No hard events and two total events occurred in patients in the lowest quartile (CAC score 0). In those patients with a scores from 1 to 4, 5 to 332, and 333 to 4607, one, ten and 19 hard events occurred, and one, 27 and 27 total events occurred, respectively. Thus, all cases of cardiac death and MI occurred in patients with a CAC score of > 0 and the annualized event rate for all cardiovascular events was 0.6% for the 76 subjects with a CAC score of 0 compared to 13.9% for the 38 subjects with a score of > 400. ($P < .001$) The authors conclude that CACS of 0 may expedite early discharge from the ED in this patient population. They also note that calcium scores "should improve the physician's ability to stratify individuals at high risk of events, to whom aggressive treatment of risk factors for coronary artery disease can be more appropriately directed, and help direct the admission or discharge of emergency room patients".

Laudon [21] reported on a series of 105 patients who received CAC scoring by EBCT within 24 hours of admission to the ED with agina-like chest pain, normal cardiac enzymes, and a normal or nondiagnostic ECG. Since previous studies have shown that the prevalence of CAC increases markedly with age, men had to be under 55 years of age and women under 65 years of age. Other cardiac testing (treadmill exercise test, radionucleotide stress test, angiography, stress echocardiography) was also performed in 100 of the patients (95%) at the discretion of the staff physician without knowledge of the calcium scores. An EBCT examination was considered positive if any coronary calcium was detected (CAC score > 0). All patients were followed up 4 months later using chart review and telephone calls. Among the 54 patients with CACS of 0, other cardiac tests were negative. (One patient's treadmill exercise test was initially read as positive, but re-read by an independent cardiologist and classified as negative; the authors classified this patient as having a negative cardiac test.) Among the 46 patients with a calcium score > 0, 14 (30%) had other cardiac tests that were positive. All patients with a calcium score of 0 were free of cardiac events during 4 months of follow-up. The authors suggest that a calcium score of 0 would allow a patient to be discharged "from the ED without further testing, with referral to his or her primary physician for outpatient evaluation." A calcium score > 0 would require further evaluation in a chest pain unit or in the hospital.

McLaughlin [106] enrolled 134 low-risk patients with the primary goal of evaluating the predictive value of a negative EBCT scan in terms of risk stratification. The population consisted of patients admitted to an emergency room with chest pain, normal or nondiagnostic ECGs, and normal cardiac enzymes. Whether the study was prospective or retrospective is not stated. All patients had calcium scores obtained within 7 days of admission and treating physicians were not told those results. End points were acute MI, percutaneous revascularization, coronary artery bypass, or sudden cardiac death as ascertained by chart review and telephone contact 30 days after the hospitalization. Among the 48 (36%) patients whose calcium score was 0, only one had a cardiac event (a cocaine abuser had an acute MI). Among the 86 patients (64%) whose calcium score was > 0, there were seven cardiac events. All but one event occurred during the index

hospitalization. The authors conclude that calcium scores should not be used for risk stratification in cocaine abusers; excluding that patient "this test was perfect in predicting which patients may be safely discharge from the emergency room." Also, "identification of this high-risk subgroup (with calcium score > 0)…may allow for more appropriate evaluation and aggressive treatment in patients admitted from the emergency room to the hospital with chest pain."

Esteves [103] evaluated the utility of CAC to predict a normal adensosine stress rubidium-82 (Rb-82) myocardial perfusion positron emission tomography (PET) in 84 patients admitted to a chest pain unit with normal or nondiagnostic ECGs and two negative sets of cardiac enzymes. All patients subsequently had adenosine stress Rb-82 myocardial perfusion PET/CT. No clinical follow-up was performed as the authors used normal PET results as a proxy for good short-term outcomes, citing its wide acceptance as "a tool used to exclude functionally significant coronary stenosis and is associated with a very low risk of short-term cardiac events." In the 34 patients with a calcium score of 0, the PET scan was normal. In the 50 patients with a calcium score > 0, myocardial perfusion defects were seen in 13 (26%). Also, left ventricular ejection fraction was generated in 72 of the patients and was normal in 30 or 31 (97%) patients with no CAC and in 37 of 41 (90%) patients with CACS > 0. The authors conclude that absence of CAC is predictive of a normal adenosine stress Rb-82 myocardial perfusion PET and that myocardial perfusion imaging probably can be safely avoided in chest pain patients with a CACS = 0, saving the patient from further unnecessary radiation exposure and increased costs.

The extent to which CACS may facilitate decision making in the emergency department is not clear from these studies.

Prediction of future events in non-emergency settings

No randomized trials which demonstrate that treatment of cases detected by the diagnostic test improves patient outcomes were found. In the absence of such studies, even though an association between CACS and future cardiovascular events has been reported, the extent to which various treatment options may have influenced the associations is unclear. It is unknown to what extent use of CACS may truly influence patient outcomes.

Three reports which evaluated the association between CACS and future "hard" events such as death or myocardial infarction were identified. Differences in outcomes measure definition, underlying CAD risk and length of follow-up were found across studies. A brief description of the studies reporting on symptomatic persons is provided below.

Three studies were conducted populations from non-emergency settings. [107-109] In general, all report that CACS above a low threshold appears to be a predictor for hard events and that a CACS = 0 or one that is "low" was associated with few such events. The risk for future events increased with increasing CACS.

Summary of cardiac events occurring at various CACS scores

Author (year)	Total cardiac events	Number of cardiac events: CACS = 0 or low range	CACS > 100	CACS > 400 or high range
Keelan (2001) N = 317	n = 22	n = 1*	n = 17	n = 7†
Schmermund (2004) N = 300	n = 40	n = 1*	n = 35	n = 21†
Kennedy (1998) N = 368	n = 13	n = 1	n = 9	n = 7

*At the low range of calcium scores, Keelan reported one hard event in the first quartile of CACS ≤ 12; and Schmermund reported one hard event in the first quartile of CACS 0-1.4.
†Keelan reported 7 hard events for CACS ≥ 621; and Schmermund reported 21 hard events for CACS > 248.

Using a calcium score cut-off of ≥ 100 compared with calcium scores < 100, one study found that 71% of hard events (cardiac death and nonfatal MI) occurred in patients in this score range, and reported a unadjusted relative risk of a hard event of 3.20 (95% CI, 1.17-8.71). [107]A second study reported that 88% of hard events (cardiac death, MI, interventional or surgical revascularization) occurred in patients with CS ≥ 100, and the unadjusted relative risk of a hard event was 12.0, (95% CI of 4.7-30.6).[109] After adjusting for calcium score and age, a relative risk 4.4 (95% CI, 1.5-12.6) was reported. The confidence intervals for these estimates reflect wide variation in the estimates.

The results of these studies should be interpreted cautiously. Definitions of events varied across studies. While all adjusted for potentially confounding risk factors measured at the time of the CACS, none reported on or evaluated the influence of treatment between the time of the CACS determination and final follow-up or event occurrence. The reported mean length of follow-up in these studies varied from 30 months to 83 months with the low range in one study of12 months. At shorter lengths of follow-up, there may not have been adequate time for observation of events.

3.4 Key Question 4: What is the evidence that CACS may perform differently in special populations (e.g. women, diabetic populations)?

Studies which provided information on CACS in symptomatic diabetic persons and women were found and several validation studies looked at the effect of age on CACS. The high prevalence of CAD among diabetic patients and lower prevalence in women compared with men should be borne in mind when interpreting the predictive values.

The small number and mixed quality of studies of women and diabetic persons do not provide sufficient evidence that CACS may perform differently in these populations.

Persons with diabetes are at higher risk of developing significant CAD compared with non-diabetic persons. Two studies examined a small number of symptomatic diabetic populations as a separate diagnostic group[70, 82] and evaluated the validity of calcium scoring for the presence of significant coronary artery disease (CAD), defined as ≥ 50% stenosis as determined by angiography. These studies are of moderate quality (LoE II-III). Further details concerning these studies may be found in the appendices.

Table 12. Patient characteristics from studies of diabetic patients

Author (year)	Demographics	Clinical information
Khaleeli (2001)	N = 168 Male: 60% Age: 58 years	• average of 2.06 nondiabetic cardiac risk factors per patient, including age, tobacco use, hypercholesterolemia, hypertension, family history • coronary artery disease: 1-vessel: n = 36 2-vessel: n = 41 3-vessel: n = 47
Hosoi (2002)	N = 101 Male: 70% Age: 64 years	• hypertension: 66% • lipidemia: 30% • medications to control diabetes: 63%

Setting the cardiac score cut-off at 0 appears to give maximum sensitivity to the presence of CAD, with a high positive predictive value also and a moderately high negative predictive value (75%-89%). Setting the cut-off at 100 maintains a high positive predictive value, with a range of 67%-77% for the sensitivity and similarly 75%-77% for the specificity of the test. A cut-off of 400 leads to high specificity, but low sensitivity, as reported by Hosoi et al.[82]

Table 13. Summary of CACS test characteristics in diabetic populations

Cut-off >0													
Author	**N**	**TP**	**TN**	**FP**	**FN**	**Sens**	**Spec**	**PPV**	**NPV**	**1-NPV**	**n CAD**	**% CAD**	**LoE**
Khaleeli (2001)	168	122	17	27	2	98%	39%	82%	89%	11%	124	74%	II
Hosoi (2002)	100	87	3	9	1	99%	25%	91%	75%	25%	88	88%	III
Cut-off >100													
Author	**N**	**TP**	**TN**	**FP**	**FN**	**Sens**	**Spec**	**PPV**	**NPV**	**1-NPV**	**n CAD**	**% CAD**	**LoE**
Khaleeli (2001)	168	95	34	10	29	77%	77%	90%	54%	46%	124	74%	II
Hosoi (2002)	100	59	9	3	29	67%	75%	95%	24%	76%	88	88%	III
Cut-off >400													
Author	**N**	**TP**	**TN**	**FP**	**FN**	**Sens**	**Spec**	**PPV**	**NPV**	**1-NPV**	**n CAD**	**% CAD**	**LoE**
Hosoi (2001)	100	43	11	1	45	49%	92%	98%	20%	80%	88	88%	III

CAD: coronary artery disease; FN: false negative.; FP: false positive; LoE: level of evidence; Sens: sensitivity; Spec: specificity; PPV: positive predictive value; NPV: negative predictive value; NR: not reported; TN: true negative; TP: true positive.

The prevalence of CAD is much higher in these studies (74%-88%) in diabetic persons compared with the validation studies in more general populations (49%- 73%) a factor to consider when interpreting the predictive values. Hosoi reported that the sensitivity and specificity were not significantly different between the diabetic and nondiabetic persons in his study.[82] Kaheeli reports that a cut off of 102 provided optimal sensitivity (77%) and specificity (77%) to detect obstructive CAD in symptomatic diabetic persons. These authors also point out that prevalence of any coronary calcium is higher in symptomatic diabetic patients (89%) than in symptomatic non-diabetic patients (73%).[70]

The pooled estimates from LoE I/II studies in this HTA are listed below for comparison;

	LoE I or II studies
Sensitivity	99 % (98%, 99%)
Specificity	35% (33%, 36%)
PPV	65% (63%, 66%)
1-NPV	5% (4%, 6%)

Three studies reported on the validity of calcium scores in male and female populations[67, 68, 111] as shown in Table 13. These were LoE II-III, and further details can be found in the appendices.

Table 14. Patient characteristics in male and female study populations

Author (year)	Demographics	Clinical information
Budoff (2002)	N = 387 Age: 58 years Female: 100%	• hypertension: 57% • diabetes: 24% • hypercholesterolemia: 48% • current tobacco use: 25% • family history of CAD: 44% • coronary artery disease: 1-vessel: n = 72 2 or more vessels: n = 88
	N = 733 Age: 56 years Male: 100%	• hypertension: 49% • diabetes: 17% • hypercholesterolemia: 33% • current tobacco use: 23% • family history of CAD: 44%
Rumberger (1995)	N = 50 Age: 56 years Female: 100%	• history of MI: 2%
	N = 89 Age: 47 years Male: 100%	• history of MI: 4%
Haberl (2001)	N = 539 Age: 60 years Female: 100%	NR
	N = 1225 Age: 56 years Male: 100%	NR

NR = not reported

A calcium score cut-off of 0 appears to give maximum sensitivity (96%-100%) and negative predictive value for diagnosis of CAD, but specificity and positive predictive value are low, for both male and female patients. A cutoff of > 100 improves the specificity of calcium scoring while sensitivity and negative predictive value are still moderately high.[68]

Examining the value of 1-NPV, a range of 4-11% of men who tested negative but actually had CAD would be missed at the CS = 0 level, whereas only 0-4% of women would be missed. Values at a cut-off of >100, are 11% of men and 18% of women,

The prevalence of CAD was lower in women (36%-47%) compared with men (53%-70%). Women present with CAD at an older age (~10 years) than men, which may account for the differences in the prevalence and predictive values. In two of the studies, men and women were of similar ages[67, 68] and the third study enrolled a relatively young population of both men and women but women were 10 years older than men on average.[111] All studies enrolled many fewer women than men.

Table 15. Summary of CACS performance in male and female populations compared with angiography

Cut-off >0													
Author	**N**	**TP**	**TN**	**FP**	**FN**	**Sens**	**Spec**	**PPV**	**NPV**	**1- NPV**	**n CAD**	**% CAD**	**LoE**
Female populations													
Budoff (2002)	387	154	130	98	5	96%	57%	61%	96%	4%	159	41%	II
Rumberger (1995)	50	18	21	11	0	100%	66%	62%	100%	0%	18	36%	III
Haberl (2001)	539	255	116	168	0	100%	41%	60%	100%	0%	255	47%	II
Male populations													
Budoff (2002)*	733	NR	NR	NR	NR	96%	46%	NR	89%	11%	512	70%	II
Rumberger (1995)	89	46	24	18	1	98%	57%	72%	95%	5%	47	53%	III
Haberl (2001)	1225	680	128	412	5	99%	24%	62%	96%	4%	685	56%	II
Cut-off >100													
Author	**N**	**TP**	**TN**	**FP**	**FN**	**Sens**	**Spec**	**PPV**	**NPV**	**NPV**	**n CAD**	**% CAD**	**LoE**
Female populations													
Haberl (2001)	539	209	216	68	46	82%	76%	75%	82%	18%	255	47%	II
Male populations													
Haberl (2001)	1225	637	405	135	48	93%	75%	83%	89%	11%	685	56%	II

CAD: coronary artery disease; FN: false negative.; FP: false positive; LoE: level of evidence; Sens: sensitivity; Spec: specificity; PPV: positive predictive value; NPV: negative predictive value; NR: not reported; TN: true negative; TP: true positive.
*Raw data was not extractible from the report; those values reported here are as listed in the report. Other numbers given in this table are as reported in that study or were calculated from the reported numbers.

Seven LoE I/II validation studies looked at the effect of age on calcium scores and diagnostic accuracy.[66-69, 72, 74, 112] Two studies which detailed results by age are summarized below. Further details on these higher quality studies are available in the appendix and other sections of this report.

Table 16. The effect of age on diagnostic validity of calcium scores.

Author (year)	Age group (years)	N	CACS cut-off	Sens	Spec	PPV	NPV	% CAD
Kajinami (1995)	< 40	16	0.0*	50%	75%	40%	82%	NR
LoE I	40 to ≤ 50	47	2.0*	86%	94%	86%	94%	NR
N = 251	50 to ≤ 60	57	2.0*	82%	75%	74%	82%	NR
	60 to ≤ 70	78	4.0*	69%	94%	97%	61%	NR
	≥ 70	53	4.0 *	78%	76%	88%	62%	NR
Lamont (2002)	< 50	27		93%	83%	88%	91%	56%
LoE II	50-60	59	> 0	98%	50%	86%	88%	76%
N = 153	> 60	67		100%	67%	91%	100%	78%

CAD: coronary artery disease; FN: false negative.; FP: false positive; LoE: level of evidence; Sens: sensitivity; Spec: specificity; PPV: positive predictive value; NPV: negative predictive value; NR: not reported; TN: true negative; TP: true positive.
*Log-transformed total coronary calcification scores. Scoring as follows in Hounsfield units (HU): 1 = 130-199 HU; 2 = 200-299 HU; 3 = 300-399 HU; 4 = 400+ HU.

The prevalence of CAD and presence of calcium increases with age. There are, however somewhat mixed results regarding the extent to which age influences test performance characteristics.

Increase in coronary calcification was found to be significantly associated with increased age, regardless of gender or presence or absence of significant stenosis.[68, 69] Kajinami et al also reported that the most prominent difference in log-transformed total coronary calcium scores between patients with and without CAD was among men aged 40 to ≥60 years and in women ≥60 years, and that patients with CAD had significantly higher calcium score values than those without CAD.[69] Sensitivity a tended to vary with age with one of the studies[72] showing an increase with age while the other, Kajinami reporting the highest degree of sensitivity among middle-age patients (40 to ≤60 years) in their study population, while specificity and NPV was found to decrease with age (< 40 years). Data from Knez suggests that the optimal trade-off between sensitivity and specificity occurs in those 50 -70 years based on ROC.[112]

Budoff et al found that whereas there was no significant difference in sensitivity for detection of CAD in comparing young and old patients at the cut-off CS > 0, a significant difference was determined in specificity for premenopausal women (age < 60 years, 65%) and postmenopausal women (42%).[67]

In other studies, the absence of any calcium had a high "predictive accuracy" for ruling out obstructive CAD in subjects ≥ 50 years of age in two studies,[66, 112] and in all age

ranges in one study.[68] Two studies reported a high rate of false negatives among younger patients.[74, 112] Knez reported that of eight patients (seven men, one woman) with a calcium score of 0 and found to have to have significant stenosis on coronary angiography, seven were < 45 years old. Similarly, Leber et al reported that three out of four patients with significant stenosis despite low calcium score were also aged < 45 years.[74]

3.5 Key Question 5: What is the evidence of cost implications and cost-effectiveness for CACS compared with other diagnostic tests?

Two full formal economic analyses describing CACS use in symptomatic patients[113, 121] were identified and were of moderate quality. One poorer quality costing study was also found.[95] All three articles evaluated the use of CACS as a triage before conventional coronary angiography (CCA). Though all included CCA alone as one comparator modality, each also included different additional noninvasive modalities and cutoff values for calcium score, which makes comparison between studies somewhat challenging.

The two full economic studies form the focus of this report with the costing study only briefly described. The most recent study [113] was conducted in Europe which may limit its generalizability to a US system since referral patterns and reimbursement policies may differ. The two studies are US-based and are at least 10 years old and did not include comprehensive measures, societal perspectives, or time horizons that would allow strong conclusions about the economic value of CACS compared with angiography or other tests.

The two full economic analyses modeled multiple hypothetical cohorts and assumed that all patients with positive or non-diagnostic findings on the initial noninvasive test would receive CCA. Cost-effectiveness is characterized by use of intermediate clinical outcomes rather than survival or quality-adjusted survival. The use of an intermediate outcome (in these studies, cost per correct diagnosis) may be a relevant clinical endpoint but (1) makes it difficult or impossible to compare with economic evaluations conducted in other clinical areas and (2) by definition does not provide information about survival as an outcome. The authors of one study state that quality adjusted life years (QALYs) are not the best measure to use for studying diagnostic interventions, but other studies have successfully used QALYs as the outcome of interest in economic evaluations in related clinical areas,[122, 123] and at least some experts support the use of QALYs even for diagnostic testing where there is a relative distance between diagnosis and health outcome.[124]

Overall, there is weak evidence (based on the two moderate quality full economic analyses) that at low to moderate pretest disease prevalence of CAD, CACS may have a lower cost per correct diagnosis than angiography alone. However, it is worth noting that CACS may not be the most cost-effective compared with other non-invasive diagnostic pathways. One author concluded that calcium scoring cannot be recommended from an economic perspective. [113] Differences in model assumptions and failure of the studies to

describe an incremental cost-effectiveness ratio for various strategies, make drawing a firm conclusion regarding cost effectiveness of CACS challenging.

- The cost of CCA per test as measured by hospital fees is considerably higher than CACS (nearly 8 times higher, according to Rumberger). However, since CCA is still required as a second test for differential diagnosis of some proportion of people receiving CACS, the value of CACS must be interpreted in the context of total number of conventional angiograms conducted.

- CAD prevalence is an important driver of the cost per correct diagnosis: The costs decreased with increasing pretest likelihood of disease.
    - In one study, modeled CACS costs per correct diagnosis decreased from €2345 to €1897 as pretest CAD likelihood increased from 30% to 40%. Between these likelihoods, CACS was more cost effective than the any of the traditional diagnostic modalities.[113]
    - In the other study, at each CACS cutoff chosen, as the prevalence of CAD increased, CACS costs decreased for all modalities. At a threshold of > 0, modeled CACS costs went from $24,703 USD when CAD prevalence was 10% to $6,329 and $4,957 respectively as prevalence increased to 50% and 70%.[121]

- Accuracy of CACS and CACS threshold appear to influence cost effectiveness.
    - Sensitivity analysis done by maximally increasing and decreasing test accuracy (not defined by authors) within the 95% confidence interval suggested that CACS was more cost effective than traditional approaches only at a pretest likelihood of 40% when accuracy was decreased but more cost effective than the alternatives when the accuracy was maximally increased at pretest likelihoods of 20%- 50%.[113]
    - At an intermediate disease prevalence of 50%, The direct costs per correct diagnosis in USD were $6,329, $5,410, $5,290 and $5,186 for CACS thresholds of > 0, 37, 80 and 168 respectively. The corresponding true positive rates at these cut offs were 96%, 90%, 84% and 72%.[121]

- The rates, health consequences and cost impact of false positives and false negatives are not clearly described in these studies for CACS or relative to other testing alternatives.

- The extent to which the models presented reflect clinical practice is unclear. CACS does not appear to function as a stand alone test. Patients with positive CACS will have additional testing to assess function (e.g. stress test) or angiography, which is the definitive anatomical test. Since the role of CACS is not clear, the extent to which clinicians do not trust CACS results may influence the use of tests that are more familiar, such as ETT. The potential impact of these practices on cost and benefit need to be considered.

- There is insufficient evidence for conclusions on the long-term cost utility of CACS compared with CCA alone or with regard to other non-invasive tests.

Economic study detail
The table below provides a summary of the perspectives and data sources used in the economic studies.

Table 17 Overview of economic studies for CACS

	Design	Perspective and Costing	Data sources and Population	Primary Strengths	Primary Limitations
Dewey 2006	Cost-effectiveness; Cost per correctly identified obstructive CAD patient in Euros German health care system 10 year time horizon	Societal perspective Direct and indirect costs Provider perspective for MSCT break even costs	Data from meta-analyses Established German reimbursement rates Hypothetical cohort of suspected CAD patients aged from 30-69 years	Sensitivity analysis done on test parameters, CCA costs and complications Costs of missed diagnosis included 5% annual discounting for complications	EBCT performance data from meta-analysis which included other diagnostic modalities and populations Details of search and primary sources not provided EBT specific sensitivity data were not detailed Price per year is not detailed
Rumberger 1999	Cost-effectiveness US system Cost per correctly diagnosed obstructive CAD	Perspective not stated, appears to be provider perspective Total direct costs Local non-Medicare fee schedule	Data from 213 patients from a previous study used for model input parameters Hypothetical cohorts of 100 pts for each of 8 possible pathways were modeled	Test and its complications included Sensitivity analysis for different prevalence and thresholds	Details of cost, prevalence for complications not provided No sensitivity analysis on cost or other model assumptions
Raggi 2000	Cost-only US system Cost per patient tested in USD	Mean reimbursement from 3 Tennessee payers Direct costs only Apparently payer perspective	N = 207 Age 50 years (± 9) range 35-67 mean pretest probability 29% (median 22%) Persons with low to intermediate pretest probability of CAD who received EBCT and ETT in random order	Sensitivity analysis for test parameters and pretest likelihood < 15%, ≥ 15% to < 85% Considered cost of additional testing	Limited sensitivity analysis Model doesn't include complications Date of costs not given

USD = United States Dollars, MSCT = multi-slice computed tomography; EBCT = electron beam computed tomography; ETT = exercise treadmill testing

The quality of these studies was variable. QHES scores were 76 and 70 (out of 100) for the full analyses and 46 for the costing study. The two full analyses evaluated the effect of varying CAD prevalence and CACS threshold on cost per correct diagnosis. One study, [DEWEY] briefly described sensitivity analyses around model assumptions and cost inputs but these were not fully described for CACS. The other did not report any additional sensitivity analysis. The sensitivity analyses should ideally permit a transparent assessment of relative contributions of inputs that drive the model results, for

example, use of tornado plots. Thus, a clear understanding of the drivers of cost-effectiveness is difficult.

The cost per correct diagnosis is provided in these studies, however, incremental cost-effectiveness ratios for one strategy versus another were not described in either of the full economic studies. Though the analyses included CCA alone as one comparator modality, each also included different additional noninvasive modalities and cutoff values for calcium score, which makes comparison between studies somewhat challenging.

The sensitivities and specificities for CACS used in the two full economic studies compared with those in the present HTA are provided below.

	Sensitivity (95% CI)	**Specificity (95% CI)**
Present HTA LoE I/II studies		
CACS >0	99% (98%, 99%)	35% (33%, 36%)
CACS ≥100	85% (84%, 86%)	77% (76%, 79%)
CACS ≥ 400	78% (70%, 86%)	83% (76%, 91%)
Dewey (Cut-off not specified)	92.3% (90.7%, 94.0%)	51.2% (47.5%-54.9%)
Rumberger		
CACS >0	95% (CI not reported)	46%(CI not reported)
CACS = 37	90%	77%
CACS = 80	84%	84%
CACS = 168	71%	90%

CI = confidence interval

Results of formal economic studies

Each of the formal economic analyses employed different modeling assumptions and diagnostic pathways of interest. A summary of the primary results of each are presented in Table 18.

Table 18. Overview of results from full economic studies

	Diagnostic pathway of interest	Comparator(s)	Relevant results	Results of sensitivity analysis	Author conclusions
Dewey 2007	CACS followed by CCA for positive or nondiagnostic results (cutoff calcium scores not presented—but used sensitivity = 92.3 and specificity = 51.2)	CCA alone; MSCT, stress MRI; exercise ECG, or stress echocardiography CACS followed by CCA for positive or non-diagnostic results	At prevalence 10%-50%, MSCT most cost effective pathway (CER= €4435 @10%, €1469 @ 50%) At prevalence 60%, MSCT and CCA alone were equally cost-effective (CER=€1345) At prevalence >70%, CCA alone was most cost-effective pathway (€1153 @70%, €807 @100%) At prevalence 30%-40%, CACS was more cost-effective than all other pathways except MSCT (CER €2345 @30%, €1897 @40) At prevalence >50%, CCA more cost effective than CACS. Prevalence at which CCA alone becomes more cost-effective than CACS > 60% (note MSCT found to be more cost-effective than CACS)	Altering sensitivity/specificity, complication rates, did not alter order of main findings Altering reimbursement rate for CCA altered prevalence at which CCA became most cost-effective (50%-80%) MSCT was only most cost-effective option at reimbursement rate ≤€260; at higher cost was overtaken by other noninvasive modalities Break-even analysis: 64 months @ 10 coronary exams per day; 23 months in higher volume facility	Up to 50% pretest likelihood of CAD, CACS is more cost-effective from societal perspective than CCA alone; above 50% prevalence CCA is most cost-effective option. (Note MSCT found to be more cost-effective than CACS at prevalence <60%) Calcium scoring cannot be recommended from an economic perspective
Rumberger 1999	CACS with EBCT followed by confirmation with CCA for positive or nondiagnostic results (cutoffs CACS > 0; CACS = 37, CACS = 80, CACS =168)	CCA alone; treadmill exercise testing alone (TMET), with 2D echocardiography (ECHO), or with thallium scintigraphy (THALLIUM) (all exercise pathways followed by confirmation with CCA for positive or nondiagnostic results)	For disease prevalence at or below 70%, the least costly and most cost-effective pathway considered was a CACS cutoff of 168 (sens = 71%, spec = 90%). At 100% prevalence, angiography alone had the lowest cost per correct diagnosis. Angiography alone cost per correct diagnosis: $35,400 @ 10% prevalence, $3540@ 100% prevalence CACS at a score 168 cost per correct diagnosis: $15,016 at 10% prevalence, $4071 at 100% prevalence. Prevalence at which CCA alone becomes more cost-effective than CACS = >70%	Results presented for varying levels of disease prevalence and CS cut-off scores; costs not varied	In population with prevalence ≤70%, CACS with EBCT minimized direct costs and maximized cost-effectiveness; at 70%-100% prevalence CCA is superior option

MSCT = multi-slice computed tomography; EBCT = electron beam computed tomography; CER = cost-effectiveness ratio; CACS = coronary artery calcium score

The most recent and detailed evaluation by Dewey takes a modified societal perspective, using a time horizon of 10 years. A decision-analytic model was used and included evaluation of six different strategies for patients presenting with stable chest pain: CACS (based on EBCT), CCTA, stress echo, stress ECG, dobutamine stress MRI and immediate CCA. The main inputs to the model were test sensitivity and specificity, prevalence of disease in the population, rate of nondiagnostic exams, and rate of complications; the outcome of interest was correct diagnosis of CAD. Costs included were reimbursement rates for examinations, subsequent tests, treatment of complications, and treatment of subsequent myocardial infarction (hospitalization and rehabilitation). Measures of productivity loss cost were also included for subsequent myocardial infarction. Sources for clinical data came largely from meta-analyses, and cost data from German reimbursement schedules and author decision. Detailed methods of data abstraction were not provided. Sensitivity analysis included varying levels of test accuracy, complication rate of angiography, and costs. They also conducted a break-even analysis from a provider perspective. Results for each strategy as a stand alone procedure were provided but incremental differences among strategies were not presented.

The authors combined benefits and costs with a cost-effectiveness analysis, using cost per correct diagnosis as the unit of analysis, which they present for all the modalities at several levels of prevalence in the population. In general, cost-effectiveness increased as the prevalence of disease increased. At disease prevalence above 60%, angiography alone was the most cost-effective option of the modalities examined; at 30% to 40% prevalence CACS was most cost effective except for MSCT; and at 10%-20% prevalence and 40%-60% prevalence MSCT was most cost-effective. The relative cost-effectiveness was not changed with sensitivity analysis, except for the varying of reimbursement cost of MSCT. The authors conclude that MSCT, not CACS, is the most cost-effective noninvasive modality at prevalence below 60%. However, comparing only CACS with angiography alone, CACS remained more cost-effective at prevalence up to 50%.

Limitations of the Dewey study deserve mention. First, the use of an intermediate outcome (correct diagnosis) may be a relevant clinical endpoint but makes it difficult or impossible to compare with economic evaluations conducted in other clinical areas. Additionally, they use a mathematical model instead of patient-level data. Finally, the generalizability of a German study to a US system is unknown, both in clinical relevance and costs. Specifically, their finding that reimbursement rates were the only variable that changed the order of cost-effectiveness of various non-invasive modalities may be significance for international comparisons.

Rumberger (1999)[121] undertook an economic evaluation of CACS with EBCT for the diagnosis of CAD, compared with angiography alone, treadmill exercise, stress echocardiography, stress thallium (followed by CCA if indicated). Using data from a published study of a Mayo Clinic cohort of men and women under evaluation for CAD for the clinical inputs, the model included sensitivity, specificity, nondiagnostic rate, rate

of complications, and non-Medicare fees for testing and for a hypothetical cohort of patients. Several cutpoints of calcium score were explored in the model: >0, 37, 80, and 168, as were several prevalence rates (10%, 20%, 50%, 70%, and 100%). Only direct costs of testing and its complications were considered.

The authors report that for disease prevalence at or below 70%, the least costly and most cost-effective pathway considered was a CACS with a calcium score cutoff of 168. Compared to angiography alone, whose cost per correct diagnosis ranged from $35,400 at 10% prevalence to $3540 at 100% prevalence, CACS at a cutoff score of 168 was $15,016 at 10% prevalence and $4071 at 100% prevalence. At 100% prevalence, angiography alone had the lowest cost per correct diagnosis.

This study is a very short-term economic evaluation of the direct costs and outcomes of CACS as a triage diagnostic strategy for CAD. It provides some useful information on the relative cost and cost per correct diagnosis of various noninvasive modalities, but the lack of a defined perspective (it appears to be a hospital perspective since only hospital fees are included), survival or quality-adjusted survival or long-term time horizon makes the usefulness of the study as an economic evaluation limited.

Raggi (2000)[95] conducted a costing study to determine the relative costs of exercise treadmill testing (ETT) with CACS as the initial test for investigation of chest pain, with the hypothesis that CACS using EBCT is effective and/or lower cost than ETT in people with low to intermediate CAD likelihood—as such, they examined two primary diagnostic algorithms: CACS or ETT followed by myocardial perfusion imaging (MPI) and CCA. The authors constructed a Bayesian model and conducted a clinical study. For the clinical study, people underwent ETT and CACS in random order. Clinical outcomes collected were treadmill score, calcium score and positive, negative, or equivocal test result. Cost of testing was based on Tennessee reimbursement rates and published literature and is presented in a cost-per-patient format.

For the Bayesian model, the authors considered the CACS and ETT diagnostic pathways, the performance of MPI and angiography as the initial test, using inputs from published literature. For sensitivity, the model varied levels of calcium score and test sensitivity/specificity, and published 95% confidence intervals for their cost estimates.

The results of the modeling study suggested that both diagnostic pathways (ETT and CACS) were lower cost than angiography alone at all levels of prevalence, with CACS as the lowest cost and with cost savings greatest at lower disease prevalence. Compared with a cost of $4800 per patient tested for angiography, the primary CACS pathway ranges from $330 at 0% prevalence, to about $1500 per patient tested at 50% prevalence, to about $2200 per patient tested for 100% prevalence (numbers extrapolated from graph). In the clinical study, the CACS pathway was found to be cost-saving, with the primary pathway costing $599 per patient tested and the ETT pathway was $1701. (The clinical study did not compare CACS to CCA alone.)

This study is a cost-only study and cannot be considered a full economic evaluation. The data presented may give some estimate of budget impact, and the strength of having clinical data to support a mathematical model is probably of some value. However, the authors do not set out to determine the value of CACS as a diagnostic intervention using a synthesis of costs and benefits. Nor does the study take a long-term perspective, so the value of CACS beyond the initial testing cost is not considered. Overall, this study offers very limited information on the cost of CACS compared to angiography and none on its cost-effectiveness.

Further economic modeling

The following are areas for which there is insufficient evidence for a comprehensive assessment of economic value. These should be considered for future economic evaluations.

- Better delineation and modeling of the costs and consequences of false positive and false negative results in the short- and long- term (e.g. presentation of patients with false negative results at a future time).
- The costs and consequences of additional or unnecessary testing should be considered. The cost of follow-up and care related to incidental findings should also be considered in modeling.
- The long-term costs and outcomes of CACS compared to CCA alone on survival.
- The long-term risks to CACS, including risks radiation exposure with CACS.
- The long-term impact of CACS compared to CCA alone from a patient perspective (eg. anxiety/reassurance, test invasiveness, out of pocket costs, cost of time spent in testing), shown to be relevant in other imaging settings.[124, 125]
- Sensitivity analysis to determine the relative impact of test characteristics, pretest disease prevalence, cost, and other relevant variables on overall cost utility of CACS.

Summary and Implications

1. CACS test characteristics: Validation and accuracy, reliability and reproducibility of CACS compared with CCA.

- The role of coronary artery calcium scoring (CACS) as a diagnostic test is not clear from the literature and there is no consensus on appropriate thresholds for determining a negative versus positive test. It is not likely to be a replacement for conventional coronary angiography (CCA) based on test performance characteristics. Some literature suggests that it might be used for triaging symptomatic patients and that CACS may reduce the use of conventional coronary angiography.

- Based on meta-analysis of LoE I/II studies comparing CACS with the reference standard of conventional coronary angiography, the overall strength of evidence is high.
    - A CACS > 0 is highly sensitive (99%, CI = 98% - 99%) for identifying the presence of obstructive CAD, however specificity was only 35% and 5% of persons (1 – negative predictive value) with a negative test would have CAD based on pooled estimates from seven studies with a total of N = 7354 patients. Approximately 35% of persons without CAD might avoid unnecessary angiography or additional tests.
    - At thresholds of ≥ 100 (5 studies) or ≥ 400 (3 studies) the sensitivity is lower (85% and 78% respectively) but specificity is improved (77% and 83%, respectively). Clear decisions may not be possible based on CACS when using these thresholds to define a positive test.

2. Safety of CACS

The primary safety concerns for CACS relate to radiation exposure and the consequences of incidental findings.

Radiation exposure

- The overall strength of evidence regarding safety is very low primarily due to uncertainties regard the cancer-related risks due to radiation exposure particularly when CACS may lead to additional tests involving radiation. On the other hand, to the extent that CACS has the potential decrease the use of conventional angiography in some patients, overall radiation exposure would be reduced.
- To date, no large-scale epidemiologic studies evaluating cancer risk associated with computed tomography (CT) in general have been published.
- There is uncertainty and controversy with regard to the actual risk of low dose radiation. Quantification of risk specific to CACS for an individual patient is not possible.
- A typical effective dose for CACS is estimated to be 3 mSV (reported range 0.7 - 12 mSv) when retrospective and prospective gating are considered together. Exposure is less when scans are prospectively gated. Some experts consider the potential for harm from radiation exposure to be clinically significant particularly given that patients may be likely to have additional tests using radiation.
- A recent simulation estimating radiation dose and cancer risk suggests that a single scan for CACS may increase lifetime cancer risk. For a single screen at 55 years of age, based on a median effective dose of 2.3 mSv, site-specific estimates for lifetime risk of radiation induced cancer suggest that most cases would be lung cancer (6/100,000 in men, 14/100,000 in women) or breast cancer (4/100,000 in women).
- Decision making between physician and patient should involve a discussion of the potential risks and benefits of CACS (and subsequent testing). Final determination of net benefit for a given clinical scenario reflects the values and judgments of the persons making the decisions.
- The extent to which CACS is an adjunct to coronary CT angiography may increase radiation exposure compared with that for CACS alone.

Consequences of Incidental findings

- The overall strength of evidence is very low.
- Data from two studies suggests that 7%-10% of symptomatic persons will have incidental findings during a CT scan for calcium scoring that require further diagnostic testing and a small percent, 1.2%, will require therapeutic intervention.
- There may be benefits to early detection and treatment of the small percentage of significant pathology found incidentally, however, there is no evidence from these studies that early detection prompted more effective treatment or enhanced patient outcomes.
- The follow-up of less serious findings may create patient anxiety in addition to exposing them to the inconvenience, costs and risks of additional testing.

3. Influence on clinical decision making and patient outcomes

- There is an association between CACS and future events: Patients with higher CACS may experience more cardiac events (e.g. myocardial infarction, revascularization, death) and those with no calcium or low scores may be less likely to have future events. The extent to which CACS truly influences outcomes is unclear, however, since its impact on clinical decision making and treatment is not described.
- Overall, the evidence is low that CACS facilitates clinical decision making. While there are a number of studies describing the potential role of CACS as a triage tool for ruling out CAD and identifying those who should have additional testing, none of the studies included a comparison group. If CACS was a perfectly sensitive test, there were no false negatives and some degree of specificity, the benefit of doing CACS as a first test for triage could be estimated in the absence of an explicit comparison group. Without this or a comparison group, it is difficult to assess the incremental benefit of CACS in clinical decision making.

4. Special populations

- Two moderate quality validation studies in symptomatic diabetic patients suggest that the sensitivity (98-99%) and specificity (25%-39%) of CACS for the detection of any calcium is similar to that for general populations from the meta-analysis of LoE I/II studies but that a higher percent (11%-25%) of persons (1 – negative predictive value) with a negative test would have CAD. The overall strength of evidence is very low.
- Three moderate quality (LoE II/III) studies described performance characteristics for men and women separately. At a CACS >0, the sensitivities for both groups were 96%-100%. Specificities for women ranged for 41%-66% and those for men 24%-57%, some what lower. A higher percent (4% - 11%) of men (1 – negative predictive value) with a negative test would have CAD compared with women (0%-4%). The prevalence of CAD was lower in women (36%-47%) compared with men (53%-70%) Women present with CAD at an older age (~10 years) than men, which may account for the differences.

- Seven LoE I/II studies explored the relationship of age with test performance characteristics. The prevalence of CAD and presence of calcium increases with age. There are, however somewhat mixed results regarding the extent to which age influences test performance characteristics. While some studies suggest that sensitivity and predictive values go up with increasing age, others suggest that the best sensitivity and specificity may be in middle aged patients (40 – 60 years). The overall strength of evidence for studies with regard to age is moderate.

5. Economic implications

- Two full economic studies and one costing evaluate CACS as a stand-alone test compared with conventional angiography.
- The two moderate quality full economic studies suggest that at a disease prevalence of up to 70%, CACS may be more cost effective than conventional angiography, however incremental cost effectiveness is not described.
- Disease prevalence and CACS score cut-off (and corresponding sensitivity and specificity) appear to influence overall cost-effectiveness.
- Models did not include evaluation of incidental findings and the influence of false-negative and false-positive tests is not clear.
- CACS does not appear to function as a stand-alone test in clinical practice. The potential impact of additional testing done in clinical practice needs to be considered and modeled.
- There is insufficient evidence for conclusions on the long-term cost utility of CACS compared with CCA alone or with regard to other non-invasive tests.

Table 2. Overall Strength of Evidence (SoE) Criteria

SoE	Description	Further Research Impact	Domain Criterion Met		
			Quality	Quantity	Consistency
1	High	Very unlikely to change confidence in effect estimate	+	+	+
2	Moderate	Likely to have an important impact on confidence in estimate and *may* change the estimate	+	-	+
			+	+	-
3	Low	Very likely to have an important impact on confidence in estimate and *likely* to change the estimate	+	-	-
			-	+	+
4	Very Low	Any effect estimate is uncertain	-	+	-
			-	-	+
			-	-	-

Table 2. Summary of findings and overall strength of evidence

Outcome	Strength of Evidence	Results
Key Question 1: Evidence regarding test characteristics and reliability		
Validity of test	1	The role of CACS as a stand alone diagnostic test is not clear. There is no consensus on threshold. Based on meta analysis of LoE I/II studies • A CACS > 0 is highly sensitive (99% , CI = 98% - 99%) for identifying the presence of obstructive CAD, however 5% of persons (1 – negative predictive value) with a negative test would have CAD • At thresholds of ≥ 100 (5 studies) or ≥ 400 (3 studies) the sensitivity is lower (85% and 78% respectively) but specificity is improved (77% and 83%, respectively)
Reliability of test	1	• The reliability of CACS (based on Agaston method) appears to be moderate to high based on 3 small LoE II studies and descriptions in it two validation studies
Key Question 2: Evidence regarding safety		
Radiation	4	• While simulation and modeling of the effects of radiation exposure provide important insights into the possible changes in risks, the true attributable risk from radiation-based diagnostic tests may be difficult to determine. • Radiation exposure may be reduced to the extent that CACS use avoids doing angiography. On the other hand, exposures may be increased to the extent that positive CACS results in additional testing. • A typical effective dose for CACS is estimated to be 3mSV (reported range 1-12mSv). CACS results may lead to additional testing which involves radiation. • In a recently published simulation based on a median effective dose of 2.3 mSv, site-specific estimates for life-time risk of radiation-induced cancer suggest that most cases would be lung cancer (6/100,000 in men, 14/100,000 in women) or breast cancer (4/100,000 in women). • Decision making should include discussion of the potential for such risks.
Incidental findings	4	• 7%-10% of symptomatic persons will have incidental findings during a CT scan for calcium scoring that require further diagnostic testing and a small percent, 1.2%, will require therapeutic intervention based on two studies in symptomatic persons.
Key Question 3: Evidence regarding clinical decision making and patient outcomes		
Triage in emergency department	3	• Five studies suggest that a CACS = 0 may allow discharge of patients with suspected CAD. These studies, however vary in quality. None employed a comparison group and are considered case series.
Triage in other clinical settings	4	• One study reported that referral to conventional angiography increased with increasing CACS. No comparison group was employed.
Prediction of future events	3	• While 3 studies suggest that CACS is a predictor of future cardiac events, none evaluate the role of therapeutic interventions which may influence the occurrence of such events.
Key Question 4: Evidence regarding performance in special populations		
Diabetes	4	• Sensitivity (98-99%) and specificity (25%-39%) of CACS for the detection of any calcium is similar to that for general populations from the meta-analysis of LoE I/II studies but a higher percent (11%-25%) of persons (1 – negative predictive value) with a negative test would have CAD based on two moderate quality studies.
Gender	3	• Three studies evaluated CACS characteristics in women vs. men. Sensitivities were similar for both groups at CACS > 0. Specificities for women ranged for 41%-66% and those for men 24%-57%, some what lower. • A higher percent (4% - 11%) of men (1 – negative predictive value) with a negative test would have CAD compared with women (0%-4%)., however, the prevalence of CAD was lower in women (36%-47%) compared with men (53%-70%) • Women present with CAD at an older age (~10 years) than men, which may account for the differences

Age	**2**	• Seven LoE I/II validation studies evaluated the influence of age on CACS. In general, the prevalence of coronary artery calcium increases with age. • There are conflicting results regarding test performance at various ages.
Key Question 5: Evidence regarding cost-effectiveness		
	4	• Two moderate quality studies suggest that at a disease prevalence of up to 70%, CACS may be more cost effective than conventional angiography, however incremental cost effectiveness is not described. • Cost-effectiveness is influenced by disease prevalence and CACS score cut-off (and corresponding sensitivity and specificity) • The influence of additional testing to reflect clinical practice needs to be more fully considered. • The influence of false-negative and false positive results is unclear and models did not consider follow-up of incidental findings. • There is insufficient evidence for conclusions on the long-term cost utility of CACS compared with CCA alone or with regard to other non-invasive tests.

References

1. Greenland P, Bonow RO, Brundage BH, et al. ACCF/AHA 2007 clinical expert consensus document on coronary artery calcium scoring by computed tomography in global cardiovascular risk assessment and in evaluation of patients with chest pain: a report of the American College of Cardiology Foundation Clinical Expert Consensus Task Force (ACCF/AHA Writing Committee to Update the 2000 Expert Consensus Document on Electron Beam Computed Tomography) developed in collaboration with the Society of Atherosclerosis Imaging and Prevention and the Society of Cardiovascular Computed Tomography. *J Am Coll Cardiol.* Jan 23 2007;49(3):378-402.

2. Budoff MJ, Achenbach S, Blumenthal RS, et al. Assessment of coronary artery disease by cardiac computed tomography: a scientific statement from the American Heart Association Committee on Cardiovascular Imaging and Intervention, Council on Cardiovascular Radiology and Intervention, and Committee on Cardiac Imaging, Council on Clinical Cardiology. *Circulation.* Oct 17 2006;114(16):1761-1791.

3. Greenland P, Bonow RO. How low-risk is a coronary calcium score of zero? The importance of conditional probability. *Circulation.* Apr 1 2008;117(13):1627-1629.

4. Schenker MP, Dorbala S, Hong EC, et al. Interrelation of coronary calcification, myocardial ischemia, and outcomes in patients with intermediate likelihood of coronary artery disease: a combined positron emission tomography/computed tomography study. *Circulation.* Apr 1 2008;117(13):1693-1700.

5. Kim KP, Einstein AJ, Berrington de Gonzalez A. Coronary artery calcification screening: estimated radiation dose and cancer risk. *Arch Intern Med.* Jul 13 2009;169(13):1188-1194.

6. Lloyd-Jones D, Adams R, Carnethon M, et al. Heart disease and stroke statistics--2009 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation.* Jan 27 2009;119(3):e21-181.

7. Kung HC, Hoyert DL, Xu J, Murphy SL. Deaths: final data for 2005. *Natl Vital Stat Rep.* Apr 24 2008;56(10):1-120.

8. Mathers C, Boerma T, Ma Fat D. *The global burden of disease: 2004 update.* Geneva 27, Switzerland: World Health Organization (WHO); 2008. ISBN 978 92 4 156371 0.

9. Thom T, Haase N, Rosamond W, et al. Heart disease and stroke statistics--2006 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation.* Feb 14 2006;113(6):e85-151.

10. Schoen F. Chapter 13, The Heart. In: Cotran RS KV, Collins T, ed. *Robbins Pathophysiologic Basis of Disease*. sixth ed. Philadelphia, London, Toronto, Montreal, Sydney, Tokyo: W. B. Saunders Company; 1999:543-599.

11. Gianrossi R, Detrano R, Mulvihill D, et al. Exercise-induced ST depression in the diagnosis of coronary artery disease. A meta-analysis. *Circulation.* Jul 1989;80(1):87-98.

12. Wexler L, Brundage B, Crouse J, et al. Coronary artery calcification: pathophysiology, epidemiology, imaging methods, and clinical implications. A

statement for health professionals from the American Heart Association. Writing Group. *Circulation.* Sep 1 1996;94(5):1175-1192.

13. O'Rourke RA, Brundage BH, Froelicher VF, et al. American College of Cardiology/American Heart Association Expert Consensus Document on electron-beam computed tomography for the diagnosis and prognosis of coronary artery disease. *J Am Coll Cardiol.* Jul 2000;36(1):326-340.

14. Baumgart D, Schmermund A, Goerge G, et al. Comparison of electron beam computed tomography with intracoronary ultrasound and coronary angiography for detection of coronary atherosclerosis. *J Am Coll Cardiol.* Jul 1997;30(1):57-64.

15. Mintz GS, Pichard AD, Popma JJ, et al. Determinants and correlates of target lesion calcium in coronary artery disease: a clinical, angiographic and intravascular ultrasound study. *J Am Coll Cardiol.* Feb 1997;29(2):268-274.

16. Hoff JA, Chomka EV, Krainik AJ, Daviglus M, Rich S, Kondos GT. Age and gender distributions of coronary artery calcium detected by electron beam tomography in 35,246 adults. *Am J Cardiol.* Jun 15 2001;87(12):1335-1339.

17. Uretsky BF, Rifkin RD, Sharma SC, Reddy PS. Value of fluoroscopy in the detection of coronary stenosis: influence of age, sex, and number of vessels calcified on diagnostic efficacy. *Am Heart J.* Feb 1988;115(2):323-333.

18. Sangiorgi G, Rumberger JA, Severson A, et al. Arterial calcification and not lumen stenosis is highly correlated with atherosclerotic plaque burden in humans: a histologic study of 723 coronary artery segments using nondecalcifying methodology. *J Am Coll Cardiol.* Jan 1998;31(1):126-133.

19. Simons DB, Schwartz RS, Edwards WD, Sheedy PF, Breen JF, Rumberger JA. Noninvasive definition of anatomic coronary artery disease by ultrafast computed tomographic scanning: a quantitative pathologic comparison study. *J Am Coll Cardiol.* Nov 1 1992;20(5):1118-1126.

20. Lee JK, Sagel SS, Stanley RJ, Heiken JP. *Computed Body Tomography with MRI Correlation.* Vol 1. 4 ed: Lippincott, Williams and Wilkins; 2003.

21. Laudon DA, Vukov LF, Breen JF, Rumberger JA, Wollan PC, Sheedy PF, 2nd. Use of electron-beam computed tomography in the evaluation of chest pain patients in the emergency department. *Ann Emerg Med.* Jan 1999;33(1):15-21.

22. Thomton FJ, Paulson EK, Yoshizumi TT, Frush DP, Nelson RC. Single versus multi-detector row CT: comparison of radiation doses and dose profiles. *Acad Radiol.* Apr 2003;10(4):379-385.

23. Snow V, Barry P, Fihn SD, et al. Evaluation of primary care patients with chronic stable angina: guidelines from the American College of Physicians. *Ann Intern Med.* Jul 6 2004;141(1):57-64.

24. Agatston AS, Janowitz WR, Hildner FJ, Zusmer NR, Viamonte M, Jr., Detrano R. Quantification of coronary artery calcium using ultrafast computed tomography. *J Am Coll Cardiol.* Mar 15 1990;15(4):827-832.

25. Hoffmann U, Brady TJ, Muller J. Cardiology patient page. Use of new imaging techniques to screen for coronary artery disease. *Circulation.* Aug 26 2003;108(8):e50-53.

26. Callister TQ, Cooil B, Raya SP, Lippolis NJ, Russo DJ, Raggi P. Coronary artery disease: improved reproducibility of calcium scoring with an electron-beam CT volumetric method. *Radiology.* Sep 1998;208(3):807-814.

27. McCollough CH, Ulzheimer S, Halliburton SS, Shanneik K, White RD, Kalender WA. Coronary artery calcium: a multi-institutional, multimanufacturer international standard for quantification at cardiac CT. *Radiology.* May 2007;243(2):527-538.

28. Pelberg R, Mazur W. *Cardiac CT Angiography Manual*. illustrated ed: Springer 2007.

29. Mettler FA, Jr., Thomadsen BR, Bhargavan M, et al. Medical radiation exposure in the U.S. in 2006: preliminary results. *Health Phys.* Nov 2008;95(5):502-507.

30. Mettler FA, W. H, Yoshizumi TT, Mahesh M. Effective doses in radiology and diagnostic nuclear medicine: A catalog. *Radiology.* 2008;248(1):254-263.

31. Morin RL, Gerber TC, McCollough CH. Radiation dose in computed tomography of the heart. *Circulation.* Feb 18 2003;107(6):917-922.

32. Shope TB, Gagne RM, Johnson GC. A method for describing the doses delivered by transmission x-ray computed tomography. *Med Phys.* Jul-Aug 1981;8(4):488-495.

33. Gerber TC, Carr JJ, Arai AE, et al. Ionizing radiation in cardiac imaging: a science advisory from the American Heart Association Committee on Cardiac Imaging of the Council on Clinical Cardiology and Committee on Cardiovascular Imaging and Intervention of the Council on Cardiovascular Radiology and Intervention. *Circulation.* Feb 24 2009;119(7):1056-1065.

34. Brenner DJ, Hall EJ. Computed tomography--an increasing source of radiation exposure. *N Engl J Med.* Nov 29 2007;357(22):2277-2284.

35. Horton KM, Post WS, Blumenthal RS, Fishman EK. Prevalence of significant noncardiac findings on electron-beam computed tomography coronary artery calcium screening examinations. *Circulation.* Jul 30 2002;106(5):532-534.

36. Elgin EE, O'Malley PG, Feuerstein I, Taylor AJ. Frequency and severity of "incidentalomas" encountered during electron beam computed tomography for coronary calcium in middle-aged army personnel. *Am J Cardiol.* Sep 1 2002;90(5):543-545.

37. Achenbach S, Daniel WG, Moshage W. Recommendations for standardization of EBT and MSCT scanning. *Herz.* Jun 2001;26(4):273-277.

38. Matchar DB, Mark DB, Patel M, et al. Non-Invasive Imaging for Coronary Artery Disease. Vol Agency for Healthcare Research and Quality (AHRQ), Rockville MD. October 3, 2006. Duke Evidence-based Practice Center (EPC), Contract No. 290-02-0025.

39. Ontario Health Technology Advisory Committee (OHTAC). OHTAC Recommendation: Multidetector Computed Tomography for Coronary Artery Disease Screening in Asymptomatic Populations. May 2007.

40. BlueCross BlueShield. *Contrast-Enchanced Cardiac Computed Tomographic Angiography in the Diagnosis of Coronary Artery Stenosis or for Evaluation of Acute Chest Pain*: Technology Assessment Program; August 2006.

41. Berry E, Kelly S, Hutton J, Harris KM, Roderick P, Boyce JC. *A systematic literature review of spiral and electron beam computed tomography: with*

particular reference to clinical applications in hepatic lesions, pulmonary embolus and coronary artery disease. Health Technol Assess: NHS R&D HTA Programme.

42. Murtagh J, Warburton RN, Foerster V, et al. *CT and MRI for selected clinical disorders: a systematic review of economic evaluations.* Canadian Agency for Drugs and Technologies in Health (CADTH). August 2006.

43. The National Horizon Scanning (NHS) Centre Technology Briefing. *Computed tomography (CT) screening in risk assessment for coronary artery disease. December 2006.* University of Birmingham, Edgbaston, Birmingham, B15 2TT, England: Department of Public Health and Epidemiology.

44. Van Brabandt H, Camberlin C, I C. *64-Slice computed tomography imaging of coronary arteries in patients suspected for coronary artery disease. Health Technology Assessment (HTA)*: Brussels: Belgian Healthcare Knowledge Center (KCE); 2008. Kce Reports 82 C (D/2008/10.273/42).

45. Hall K. . *What is the prognostic value of calcium scoring in screening asymptomatic populations for cardiovascular disease? NZHTA Report 2003*: New Zealand Health Technology Assessment (NZHTA), ISBN 1-877235-46-6.

46. Waugh N, Black C, Walker S, McIntyre L, Cummins E, G H. *The effectiveness and cost-effectiveness of computed tomography screening for coronary artery disease: systematic review. Health Tech Assess*: NHS R&D HTA Programme.

47. Grady D, Chaput L, Kristof M. Results of Systematic Review of Research on Diagnosis and Treatment of Coronary Heart Disease in Women. Evidence Report/Technology Assessment No. 80. Vol AHRQ Publication No. 03-0035. Rockville, MD: Agency for Healthcare Research and Quality. May 2003. (Prepared by the University of California, San Francisco-Stanford Evidence-based Practice Center under Contract No 290-97-0013.); 2003.

48. Tice T. Utility of Coronary Artery Calcium Measurement in Cardiovascular Disease. San Francisco: California Technology Assessment Forum (CTAF). February 16, 2005.

49. Institute for Clinical Systems Improvement. Electron-Beam and Helical Computed Tomography for Coronary Artery Disease: Technology Assessment Committee. #34 Update. May 2004; 2004.

50. Centers for Medicare and Medicaid (CMS). Pub 100-03 Medicare National Coverage Determinations: Cardiac Computed Tomographic Angiography (CTA). 6/27/2008. Available at: http://www.cms.hhs.gov/transmittals/downloads/R85NCD.pdf.

51. Centers for Medicare and Medicaid Services (CMS) Regional Coverage Article. Article for Coding Guidelines-Coronary Calcium Scoring-Multidetector Computed Tomography for the Heart and Great Vessels (A45280). *Noridian Administrative Services, LLC.* Available at: http://www.cms.hhs.gov/mcd/viewarticle_pdf.asp?article_id=45280&article_version=4&contractor_id=115.

52. Aetna. Clinical Policy Bulletin: Cardiac CT, Coronary CT Angiography and Calcium Scoring. 5/06/09. Available at: http://www.aetna.com/cpb/medical/data/200_299/0228.html.

53. CIGNA HealthCare. Electron Beam Computed Tomography (EBCT) and Multidetector Computed Tomography for Coronary Artery Calcification. 12/15/2008. Available at: http://webdav.specri.com/web/documents/medicalwriters/TECHNOLOGY%20ASSESSMENT/Spectrum%20HTAs/Coronary%20Artery%20Calcium%20Scoring/Payers/Cigna%20EBCT%20and%20MDCT%20for%20coronary%20artery%20calcification%20coverage%20policy.pdf.

54. Department of Veteran Affairs Health Administration Center (CHAMPVA). CT (Computed Tomography). 2/29/08. Available at: http://www.va.gov/hac/forbeneficiaries/champva/policymanual/champva/chapter2/1c2s26-3.htm.

55. Regence. Computed Tomography to Detect Coronary Artery Calcification. 10/14/08. Available at: http://blue.regence.com/trgmedpol/radiology/rad06.html.

56. BlueCross BlueShield of North Carolina. Computed Tomography to Detect Coronary Artery Calcification. 3/2008. Available at: http://webdav.specri.com/web/documents/medicalwriters/TECHNOLOGY%20ASSESSMENT/Spectrum%20HTAs/Coronary%20Artery%20Calcium%20Scoring/Payers/BCBSNC%20policy.pdf.

57. Phillips B, Ball C, Sackett D. Levels of evidence and grades of recommendation. Available at: Available at: http://www.cebm.net/levels_of_evidence.asp.

58. Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. *Bmj*. Jun 19 2004;328(7454):1490.

59. West S, King V, Carey TS, et.al. Systems to Rate the Strength of Scientific Evidence. Evidence Report/Technology Assessment No. 47 (Prepared by the Research Triangle Institute-University of North Carolina Evidence-based Practice Center, Contract No. 290-97-0011): Agency for Healthcare Research and Quality, Rockville, MD; 2002.

60. Ofman JJ, Sullivan SD, Neumann PJ, et al. Examining the value and quality of health economic analyses: implications of utilizing the QHES. *J Manag Care Pharm.* Jan-Feb 2003;9(1):53-61.

61. Dendukuri N, Chiu K, Brophy JM. Validity of electron beam computed tomography for coronary artery disease: asystematic review and meta-analysis. *BMC Med.* 2007;5:35.

62. Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A. Meta-Disc: a software for meta-analysis of test accuracy data. *BMC Medical Research Methodology.* 2006;6(31).

63. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med.* Jul 30 1993;12(14):1293-1316.

64. Heijenbrok-Kal MH, Fleischmann KE, Hunink MG. Stress echocardiography, stress single-photon-emission computed tomography and electron beam computed tomography for the assessment of coronary artery disease: a meta-analysis of diagnostic performance. *Am Heart J.* Sep 2007;154(3):415-423.

65. Nallamothu BK, Saint S, Bielak LF, et al. Electron-beam computed tomography in the diagnosis of coronary artery disease: a meta-analysis. *Arch Intern Med.* Mar 26 2001;161(6):833-838.

66. Becker A, Leber A, White CW, Becker C, Reiser MF, Knez A. Multislice computed tomography for determination of coronary artery disease in a symptomatic patient population. *Int J Cardiovasc Imaging.* Jun 2007;23(3):361-367.

67. Budoff MJ, Diamond GA, Raggi P, et al. Continuous probabilistic prediction of angiographically significant coronary artery disease using electron beam tomography. *Circulation.* Apr 16 2002;105(15):1791-1796.

68. Haberl R, Becker A, Leber A, et al. Correlation of coronary calcification and angiographically documented stenoses in patients with suspected coronary artery disease: results of 1,764 patients. *J Am Coll Cardiol.* Feb 2001;37(2):451-457.

69. Kajinami K, Seki H, Takekoshi N, Mabuchi H. Noninvasive prediction of coronary atherosclerosis by quantification of coronary artery calcification using electron beam computed tomography: comparison with electrocardiographic and thallium exercise stress test results. *J Am Coll Cardiol.* Nov 1 1995;26(5):1209-1221.

70. Khaleeli E, Peters SR, Bobrowsky K, Oudiz RJ, Ko JY, Budoff MJ. Diabetes and the associated incidence of subclinical atherosclerosis and coronary artery disease: Implications for management. *Am Heart J.* Apr 2001;141(4):637-644.

71. Kwok B, Lim YT. Electron-beam computed tomography for symptomatic coronary disease. *Asian Cardiovascular and Thoracic Annals* 2000;8(1):46-49.

72. Lamont DH, Budoff MJ, Shavelle DM, Shavelle R, Brundage BH, Hagar JM. Coronary calcium scanning adds incremental value to patients with positive stress tests. *Am Heart J.* May 2002;143(5):861-867.

73. Lau GT, Ridley LJ, Schieb MC, et al. Coronary artery stenoses: detection with calcium scoring, CT angiography, and both methods combined. *Radiology.* May 2005;235(2):415-422.

74. Leber AW, Knez A, Mukherjee R, et al. Usefulness of calcium scoring using electron beam computed tomography and noninvasive coronary angiography in patients with suspected coronary artery disease. *Am J Cardiol.* Aug 1 2001;88(3):219-223.

75. Leschka S, Scheffel H, Desbiolles L, et al. Combining dual-source computed tomography coronary angiography and calcium scoring: added value for the assessment of coronary artery disease. *Heart.* Sep 2008;94(9):1154-1161.

76. Nixdorff U, Kufner C, Achenbach S, et al. Head-to-head comparison of dobutamine stress echocardiography and cardiac computed tomography for the detection of significant coronary artery disease. *Cardiology.* 2008;110(2):81-86.

77. Budoff MJ, Georgiou D, Brody A, et al. Ultrafast computed tomography as a diagnostic modality in the detection of coronary artery disease: a multicenter study. *Circulation.* Mar 1 1996;93(5):898-904.

78. Chen LC, Ding PY, Chen JW, et al. Coronary artery calcium determined by electron beam computed tomography for predicting angiographic coronary artery disease in moderate- to high-risk Chinese patients. *Cardiology.* 2001;95(4):183-189.

79. Fallavollita JA, Brody AS, Bunnell IL, Kumar K, Canty JM, Jr. Fast computed tomography detection of coronary calcification in the diagnosis of coronary artery

disease. Comparison with angiography in patients < 50 years old. *Circulation.* Jan 1994;89(1):285-290.

80. Guerci AD, Spadaro LA, Goodman KJ, et al. Comparison of electron beam computed tomography scanning and conventional risk factor assessment for the prediction of angiographic coronary artery disease. *J Am Coll Cardiol.* Sep 1998;32(3):673-679.

81. Herzog C, Britten M, Balzer JO, et al. Multidetector-row cardiac CT: diagnostic value of calcium scoring and CT coronary angiography in patients with symptomatic, but atypical, chest pain. *Eur Radiol.* Feb 2004;14(2):169-177.

82. Hosoi M, Sato T, Yamagami K, et al. Impact of diabetes on coronary stenosis and coronary artery calcification detected by electron-beam computed tomography in symptomatic patients. *Diabetes Care.* Apr 2002;25(4):696-701.

83. Shavelle DM, Budoff MJ, LaMont DH, Shavelle RM, Kennedy JM, Brundage BH. Exercise testing and electron beam computed tomography in the evaluation of coronary artery disease. *J Am Coll Cardiol.* Jul 2000;36(1):32-38.

84. Tanenbaum SR, Kondos GT, Veselik KE, Prendergast MR, Brundage BH, Chomka EV. Detection of calcific deposits in coronary arteries by ultrafast computed tomography and correlation with angiography. *Am J Cardiol.* Apr 1 1989;63(12):870-872.

85. Bielak LF, Rumberger JA, Sheedy PF, 2nd, Schwartz RS, Peyser PA. Probabilistic model for prediction of angiographically defined obstructive coronary artery disease using electron beam computed tomography calcium score strata. *Circulation.* Jul 25 2000;102(4):380-385.

86. Broderick LS, Shemesh J, Wilensky RL, et al. Measurement of coronary artery calcium with dual-slice helical CT compared with coronary angiography: evaluation of CT scoring methods, interobserver variations, and reproducibility. *AJR Am J Roentgenol.* Aug 1996;167(2):439-444.

87. Haberl R, Tittus J, Bohme E, et al. Multislice spiral computed tomographic angiography of coronary arteries in patients with suspected coronary artery disease: an effective filter before catheter angiography? *Am Heart J.* Jun 2005;149(6):1112-1119.

88. Konieczynska M, Tracz W, Pasowicz M, Przewlocki T. Use of coronary calcium score in the assessment of atherosclerotic lesions in coronary arteries. *Kardiol Pol.* Oct 2006;64(10):1073-1079; discussion 1080-1071.

89. Rumberger JA, Sheedy PF, Breen JF, Schwartz RS. Electron beam computed tomographic coronary calcium score cutpoints and severity of associated angiographic lumen stenosis. *J Am Coll Cardiol.* Jun 1997;29(7):1542-1548.

90. Yao Z, Liu XJ, Shi RF, et al. A comparison of 99Tcm-MIBI myocardial SPET and electron beam computed tomography in the assessment of coronary artery disease in two different age groups. *Nucl Med Commun.* Jan 2000;21(1):43-48.

91. Yao ZM, Li W, Qu WY, Zhou C, He Q, Ji FS. Comparison of (99)mTc-methoxyisobutylisonitrile myocardial single-photon emission computed tomography and electron beam computed tomography for detecting coronary artery disease in patients with no myocardial infarction. *Chin Med J (Engl).* May 2004;117(5):700-705.

92. Pepe M. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York: Oxford University Press; 2003.

93. Janssen CH, Kuijpers D, Vliegenthart R, et al. Coronary artery calcification score by multislice computed tomography predicts the outcome of dobutamine cardiovascular magnetic resonance imaging. *Eur Radiol.* Jun 2005;15(6):1128-1134.

94. Nishida C, Okajima K, Kudo T, Yamamoto T, Hattori R, Nishimura Y. The relationship between coronary artery calcification detected by non-gated multi-detector CT in patients with suspected ischemic heart disease and myocardial ischemia detected by thallium exercise stress testing. *Ann Nucl Med.* Dec 2005;19(8):647-653.

95. Raggi P, Callister TQ, Cooil B, Russo DJ, Lippolis NJ, Patterson RE. Evaluation of chest pain in patients with low to intermediate pretest probability of coronary artery disease by electron beam computed tomography. *Am J Cardiol.* Feb 1 2000;85(3):283-288.

96. Mohlenkamp S, Behrenbeck TR, Pump H, et al. Reproducibility of two coronary calcium quantification algorithms in patients with different degrees of calcification. *Int J Cardiovasc Imaging.* Apr 2001;17(2):133-142; discussion 143.

97. Serafin Z, Laskowska K, Marzec M, Sinjab TA, Lasek W, Wlodarczyk Z. Coronary artery calcium distribution and interscan measurement variability in end-stage renal and coronary heart disease patients. *Acta Radiol.* Apr 2009;50(3):288-295.

98. Jacobs PC, Mali WP, Grobbee DE, van der Graaf Y. Prevalence of incidental findings in computed tomographic screening of the chest: a systematic review. *J Comput Assist Tomogr.* Mar-Apr 2008;32(2):214-221.

99. Hunold P, Schmermund A, Seibel RM, Grönemeyer DH, Erbel R. Prevalence and clinical significance of accidental findings in electron-beam tomographic scans for coronary artery calcification. . *Eur Heart J.* . 2001;22(18):1748-1758.

100. Schragin JG, Weissfeld JL, Edmundowicz D, Strollo DC, Fuhrman CR. Non-cardiac findings on coronary electron beam computed tomography scanning. *J Thorac Imaging.* Apr 2004;19(2):82-86.

101. Law YM, Huang J, Chen K, Cheah FK, Chua T. Prevalence of significant extracoronary findings on multislice CT coronary angiography examinations and coronary artery calcium scoring examinations. *J Med Imaging Radiat Oncol.* Feb 2008;52(1):49-56.

102. Piers LH, Salachova F, Slart RH, et al. The role of coronary artery calcification score in clinical practice. *BMC Cardiovasc Disord.* 2008;8:38.

103. Esteves FP, Sanyal R, Santana CA, Shaw L, Raggi P. Potential impact of noncontrast computed tomography as gatekeeper for myocardial perfusion positron emission tomography in patients admitted to the chest pain unit. *Am J Cardiol.* Jan 15 2008;101(2):149-152.

104. Geluk CA, Dikkers R, Perik PJ, et al. Measurement of coronary calcium scores by electron beam computed tomography or exercise testing as initial diagnostic tool in low-risk patients with suspected coronary artery disease. *Eur Radiol.* Feb 2008;18(2):244-252.

105. Georgiou D, Budoff MJ, Kaufer E, Kennedy JM, Lu B, Brundage BH. Screening patients with chest pain in the emergency department using electron beam tomography: a follow-up study. *J Am Coll Cardiol.* Jul 2001;38(1):105-110.

106. McLaughlin VV, Balogh T, Rich S. Utility of electron beam computed tomography to stratify patients presenting to the emergency room with chest pain. *Am J Cardiol.* Aug 1 1999;84(3):327-328, A328.

107. Keelan PC, Bielak LF, Ashai K, et al. Long-term prognostic value of coronary calcification detected by electron-beam computed tomography in patients undergoing coronary angiography. *Circulation.* Jul 24 2001;104(4):412-417.

108. Kennedy J, Shavelle R, Wang S, Budoff M, Detrano RC. Coronary calcium and standard risk factors in symptomatic patients referred for coronary angiography. *Am Heart J.* Apr 1998;135(4):696-702.

109. Schmermund A, Stang A, Mohlenkamp S, et al. Prognostic value of electron-beam computed tomography-derived coronary calcium scores compared with clinical parameters in patients evaluated for coronary artery disease. Prognostic value of EBCT in symptomatic patients. *Z Kardiol.* Sep 2004;93(9):696-705.

110. Budoff MJ, Shokooh S, Shavelle RM, Kim HT, French WJ. Electron beam tomography and angiography: sex differences. *Am Heart J.* May 2002;143(5):877-882.

111. Rumberger JA, Sheedy PF, 3rd, Breen JF, Schwartz RS. Coronary calcium, as determined by electron beam computed tomography, and coronary disease on arteriogram. Effect of patient's sex on diagnosis. *Circulation.* Mar 1 1995;91(5):1363-1367.

112. Knez A, Becker A, Leber A, et al. Relation of coronary calcium scores by electron beam tomography to obstructive disease in 2,115 symptomatic patients. *Am J Cardiol.* May 1 2004;93(9):1150-1152.

113. Dewey M, Hamm B. Cost effectiveness of coronary angiography and calcium scoring using CT and stress MRI for diagnosis of coronary artery disease. *Eur Radiol.* May 2007;17(5):1301-1309.

114. Rumberger JA. Cost effectiveness of coronary calcification scanning using electron beam tomography in intermediate and high risk asymptomatic individuals. *J Cardiovasc Risk.* Apr 2000;7(2):113-119.

115. Agresti A. *Categorical Data Analysis (section 3.3.2).* New York: Wiley; 1990.

116. Fleischmann KE, Hunink MG, Kuntz KM, Douglas PS. Exercise echocardiography or exercise SPECT imaging? A meta-analysis of diagnostic test performance. *Jama.* Sep 9 1998;280(10):913-920.

117. US Food and Drug Administration. Radiation-emitting products. What are the radiation risks from CT? Available at: http://www.fda.gov/Radiation-EmittingProducts/RadiationEmittingProductsandProcedures/MedicalImaging/MedicalX-Rays/ucm115329.htm. Accessed July 15, 2009.

118. US Food and Drug Administration. Radiation-emitting products. Radiation quantities and units. Available at: http://www.fda.gov/Radiation-EmittingProducts/RadiationEmittingProductsandProcedures/MedicalImaging/MedicalX-Rays/ucm115335.htm. Accessed July 15, 2009.

119. Brenner DJ. Radiation risks potentially associated with low-dose CT screening of adult smokers for lung cancer. *Radiology.* May 2004;231(2):440-445.

120. Hunold P, Vogt FM, Schmermund A, et al. Radiation exposure during cardiac CT: effective doses at multi-detector row CT and electron-beam CT. *Radiology.* Jan 2003;226(1):145-152.

121. Rumberger JA, Behrenbeck T, Breen JF, Sheedy PF, 2nd. Coronary calcification by electron beam computed tomography and obstructive coronary artery disease: a model for costs and effectiveness of diagnosis as compared with conventional cardiac testing methods. *J Am Coll Cardiol.* Feb 1999;33(2):453-462.

122. Garber AM, Solomon NA. Cost-effectiveness of alternative test strategies for the diagnosis of coronary artery disease. *Ann Intern Med.* May 4 1999;130(9):719-728.

123. Sharples L, Hughes V, Crean A, et al. Cost-effectiveness of functional cardiac testing in the diagnosis and management of coronary artery disease: a randomised controlled trial. The CECaT trial. *Health Technol Assess.* Dec 2007;11(49):iii-iv, ix-115.

124. Hollingworth W. Radiology cost and outcomes studies: standard practice and emerging methods. *AJR Am J Roentgenol.* Oct 2005;185(4):833-839.

125. Mushlin AI, Kern LM, Paris M, Lambert DR, Williams G. The value of diagnostic information to patients with chest pain suggestive of coronary artery disease. *Med Decis Making.* Mar-Apr 2005;25(2):149-157.

Health Technology Clinical Committee Findings and Coverage Decision

Topic: Coronary Artery Calcium Scoring (CACS)
Meeting Date: November 20th, 2009
Final Adoption: May 14th, 2010

Number and Coverage Topic

20091120A – Coronary Artery Calcium Scoring

HTCC Coverage Determination

Cardiac Artery Calcium Scoring is a **non-covered benefit.**

HTCC Reimbursement Determination

❖ Limitations of Coverage

- Not Applicable

❖ Non-Covered Indicators

- Coronary Artery Calcium Scoring

❖ Agency Contact Information

Agency	Contact Phone Number
Labor and Industries	1-800-547-8367
Public Employees Health Plan	1-800-762-6004
Health and Recovery Services Administration	1-800-562-3022

Health Technology Background

The Coronary Artery Calcium Scoring (CACS) topic was selected and published in December 2008 to undergo an evidence review process. Heart disease is the leading cause of death and disability in US: with 700,000 deaths. The most common heart disease in the United States is coronary artery disease (CAD), which can lead to heart attack. CAD is a narrowing of one or more coronary arteries that result in an insufficient supply of oxygen to the heart muscle and is a leading cause of death in the US and developed countries. CAD may be asymptomatic or lead to chest pain (angina), heart attack, myocardial infarction (MI), or death. Cardiac related diagnostic tests include both non-invasive and invasive tests. Non-invasive tests include – stress echocardiograms: tests that compare blood flow with and without exercise and visualize the heart; single-photon emission computed tomography (SPECT), also known as nuclear stress testing or myocardial perfusion imaging; and CT angiography with or without calcium scoring using 3D imaging to visualize the heart. Invasive tests include – the "gold" standard is the conventional coronary angiography (CCA) which involves placement of a catheter and injection of contrast material into a large artery or vein, followed by 2-dimensional visualization with x-rays.

Cardiac calcium scoring uses a CT to check for the buildup of calcium in plaque on the coronary arteries. This test identifies and quantifies a marker of coronary disease (plaque), believed to detect earlier state of CAD. Cardiac calcium scoring uses CT to check for the buildup of calcium in the coronary arteries. Calcium is associated with atherosclerosis and is one marker of CAD. However, coronary calcium is not present in all atherosclerotic plaques and its relevance to risk and treatment is unclear. CACS scans the heart using CT by taking imaging "slices" of the heart. Calcium scores increase with age particularly after 50 years in men and 60 years in women.

In August 2009, the HTA posted a draft and then followed with a final report from a contracted research organization that reviewed publicly submitted information; searched, summarized, and evaluated trials, articles, and other evidence about the topic. The comprehensive, public and peer reviewed, Coronary Artery Calcium Scoring report is 94 pages, and identified a relatively large amount of literature.

An independent group of eleven clinicians who practice medicine locally meet in public to decide whether state agencies should pay for the health technology based on whether the evidence report and other presented information shows it is safe, effective and has value. The committee met on November 20th, reviewed the report, including peer and public feedback, and heard agency comments. Meeting minutes detailing the discussion are available through the HTA program or online at http://www.hta.hca.wa.gov under the committee section.

Committee Findings

Having considered the evidence based technology assessment report and the written and oral comments, the committee identified the following key factors and health outcomes, and evidence related to those health outcomes and key factors:

1. Evidence availability and technology features

The committee concludes that the best available evidence on coronary artery calcium scoring (CACS) has been collected and summarized.

- Heart disease is a prevalent and burdensome disease, and the leading cause of death in the US. Identifying which patients are at risk of major cardiac events is therefore important, but currently difficult. Symptoms of CAD (e.g. chest pain) have poor correlation to risk. Diagnostic testing can be used to help confirm or refute a suspicion of clinically significant CAD. CACS provides anatomical information (not functional) on the amount of calcium, a marker of CAD in the heart and coronary arteries.
- CACS role is unclear: it is not currently proposed or likely to be a replacement for conventional coronary angiography (CCA) based on test performance characteristics and lack of consensus about appropriate thresholds. Literature related to clinical or treatment outcomes generally focus on use for triaging symptomatic patients and that CACS may reduce the use of conventional coronary angiography.
- The clinical committee acknowledged that the population under consideration is not screening, but patients with suspected CAD. The committee discussed that this could be either asymptomatic based on history or other risk factors or symptomatic, though later concluded that most available evidence related to symptomatic patients.
- The calcium scoring process isn't automatic, experience is needed for scoring.
- A vast majority of scanners can provide a calcium score. Guidelines in 1996 provided minimum scanner requirements for resolution.

2. Is the technology safe?

The committee concludes that the comprehensive evidence reviewed is unclear in showing that calcium scoring is safe. Key factors to the committee's conclusion included:

- The committee agreed with the evidence report and current guidelines, in clinical practice, this is not a stand-alone test: it is an additional test with additional radiation and incidental findings risks. If used as triage, some individuals may not have subsequent, more invasive test, but larger group will have radiation.
- The committee agreed that there is harm in radiation exposure that is cumulative, but good evidence to quantify the risk are currently not known.
- The committee acknowledged the evidence report information regarding incidental findings, and agreed that current evidence is inconclusive.

3. Is the technology effective?

The committee concludes that the comprehensive evidence reviewed shows that Calcium Scoring is not more effective for treatment of coronary artery disease (CAD).

- The committee agreed with the evidence report and found that CACS sensitivity and reliability are high for CACS, though specificity is low and like other tests, accuracy

is affected by the disease prevalence. While accuracy and reliability are critical, they are only a first step as to whether a test is effective. The committee also agreed that there is no evidence to establish a clinically important threshold: increase in calcium does indicate disease, but the correlation to severity of stenosis is not established – which is key in a disease that is widely prevalent, where serious events occur in some, but are difficult to predict.

- In evaluating effectiveness, the most rigorous question is whether substituting this test, instead of a current diagnostic, results in better treatment and outcomes. In this case, the evidence is insufficient and current clinical practice does not support using this test alone or as a substitute.
- The other diagnostic effectiveness key question discussed by the committee is whether there is evidence that using this test as an added tool to current strategy provides a benefit (clinical or cost). The remaining analysis relate to answering this question.
- One potential use would be in ER where symptomatic patient at low to intermediate risk - could rule out disease. This use would require CACS of 0 value, so the specificity goes down, and at least a 5% group would still receive a negative test, but would have disease. One small retrospective study looked at 4 month follow up on 100 patients in ED where CACS score was taken, along with other tests and concluded that a score of 0 could permit a discharge. CACS studies did not include any RCT or higher quality observational trials to explicitly test what different clinical or treatment choices are made. Clinical expert noted that usually need a functional test to confirm.
- The committee noted national guidelines do not endorse use of Calcium scoring, though some have permissive statements for use of the test.

4. Is the technology cost-effective?

The committee concludes that the comprehensive evidence review shows no published good quality evidence on Calcium Scoring.

- Committee acknowledged the state agency costs for coronary diagnostics of nearly $7 million per year, and this would likely be an additional test and cost.
- The evidence report adequately summarized the poor cost evidence based on assumptions not current valid.
- Further, cost per correct diagnosis is a function of prevalence of disease, and CAD is highly prevalent, though the real detection issue is major adverse outcomes, not disease presence. Overall spend for reduction or prevention of negative patient outcome (here major cardiac event) is more appropriate measurement criteria.

5. Evidence about the technology's special populations, patient characteristics and adjunct treatment

- The committee agreed that no compelling evidence exists in the sub groups (diabetic, gender and age) to conclude that this test was more (or less) effective in those special populations.

6. <u>Medicare Decision and Expert Treatment Guidelines</u>

Committee reviewed and discussed the Medicare coverage decision and expert guidelines as identified and reported in the technology assessment report.

- Centers for Medicare and Medicaid Services – no national Medicare policy.
  - CMS Regional Coverage (Washington and Alaska) – the local regional CMS had determined that there is a lack of evidence of the medical necessity for quantitative evaluation of coronary artery calcium.
- Guidelines – a search of the National Guideline Clearinghouse (NGC) returned 4 potential guidelines on Calcium Scoring. The following provides a summary of the guidelines that were most relevant:
  - (1) American College of Cardiology Foundation (ACCF) – Clinical Expert Task Force – lack of evidence from studies comparing CAC measurement to alternative risk assessment techniques for moderate risk patients. No clear evidence is available indicating that additional non-invasive testing in patients with high calcium scores will result in more appropriate selection of treatment over the currently recommended preventative medical therapies. Patients with atypical cardiac symptoms may benefit from CAC testing to help exclude the presence of obstructive CAD.
  - (2) American Heart Association, 2006 – conflicting evidence and/or a divergence of opinion regarding its usefulness was found for the following indications: symptomatic patients with chest pain with equivocal or normal electrocardiograms and negative cardiac enzymes; determining the etiology; symptomatic patients in the setting of ambiguous stress tests; and asymptomatic patients with intermediate risk of CAD. Furthermore, the report stated that despite growing evidence that calcium scores are an independent predictor of CAD studies have not demonstrated improved clinical outcomes as a result of calcium score screening.
  - (3) American Heart Association, 2009 – the following are the minimum requirement which should be met in scanning for coronary artery calcium (CAC): use of an EBCT scanner or a 4-level (or greater) MDCT scanner; cardiac gating; prospective triggering for reducing radiation exposure; a gantry rotation of at least 500 ms; reconstructed slice thickness of 2.5 to 3 mm to minimize radiation in asymptomatic persons (and to provide consistency with established results); early to mid-diastolic gating; and equipment or nuclear material in cardiac imaging should be appropriately utilized to maintain patient doses as low as reasonable achievable but consistent with obtaining the desired medical information.
  - (4) American College of Radiology (ACR) Appropriateness Criteria, 2008 – for assessment of chronic chest pain in patients with low to intermediate probability of CAD: CT coronary calcium scoring received a rating of 3 (1 = least appropriate, 9 = most appropriate); a score of zero may be useful in excluding cardiac etiology; and relative radiation level is considered to be medium.
  - (5) American College of Cardiology (ACC) and American Heart Association (AHA) for the diagnosis and prognosis of CAD, 2000 – the following are a summary of interpretations and recommendations for cardiac CT scanning and CACS: a negative test (score = 0) makes the presence of atherosclerotic plaque, including unstable or vulnerable plaque, highly unlikely; a negative test is consistent with a low risk of a cardiovascular event in the next two to five years; a positive test (CAC > 0) confirms the presence of a coronary atherosclerotic plaque; the greater the amount of coronary calcium, the greater the atherosclerotic burden in men and women, irrespective of age;

and CAC measurement can improve risk predication in conventional intermediate-risk patients, and CAC plaque scanning should be considered in individuals at intermediate risk for a coronary event for clinical decision-making with regard to refinement of risk assessment.

Committee Decision

Based on the deliberations of key health outcomes, the committee decided that it had the most complete information: a comprehensive and current evidence report, public comments, input from a subject matter expert, agency and state utilization information. The committee concluded that the current evidence on Calcium Scoring demonstrates that there is insufficient evidence to cover the use of Coronary Artery Calcium Scoring (CACS). The committee considered all the evidence and gave greatest weight to the evidence it determined, based on objective factors, to be the most valid and reliable. The committee found that Calcium Scoring would be an additive test that was not supported by sufficient evidence regarding whether it is safe, cost-effective and effectively diagnoses and prevents major cardiac events thus helping patients.

Based on these findings, the committee voted 10 to 0 to not cover Calcium Scoring.

Health Technology Clinical Committee Authority

Washington State's legislature believes it is important to use a scientific based, clinician centered approach for difficult and important health care benefit decisions. Pursuant to chapter 70.14 RCW, the legislature has directed the Washington State Health Care Authority, through its Health Technology Assessment program to gather and assess the quality of the latest medical evidence using a scientific research company and take public input at all stages. Pursuant to RCW 70.14.110 a Health Technology Clinical Committee (HTCC) composed of eleven independent health care professionals reviews all the information and renders a decision at an open public meeting. The Washington State Health Technology Clinical Committee (HTCC), determines how selected health technologies are covered by several state agencies. RCW 70.14.080-140. These technologies may include medical or surgical devices and procedures, medical equipment, and diagnostic tests. HTCC bases their decisions on evidence of the technology's safety, efficacy, and cost effectiveness. Participating state agencies are required to comply with the decisions of the HTCC. Selected technologies are considered for re-review on the basis of new evidence.

National Institute for Health and Clinical Excellence

Issue date: March 2010

Chest pain of recent onset

Assessment and diagnosis of recent onset chest pain or discomfort of suspected cardiac origin

This guidance partially updates NICE technology appraisal guidance 73 (published November 2003)

NICE clinical guideline 95

Developed by the National Clinical Guideline Centre for Acute and Chronic Conditions

NICE clinical guideline 95
Chest pain of recent onset: assessment and diagnosis of recent onset chest pain or discomfort of suspected cardiac origin

Ordering information
You can download the following documents from www.nice.org.uk/guidance/CG95

- The NICE guideline (this document) – all the recommendations.
- A quick reference guide – a summary of the recommendations for healthcare professionals.
- 'Understanding NICE guidance' – a summary for patients and carers.
- The full guideline – all the recommendations, details of how they were developed, and reviews of the evidence they were based on.

For printed copies of the quick reference guide or 'Understanding NICE guidance', phone NICE publications on 0845 003 7783 or email publications@nice.org.uk and quote:

- N2113 (quick reference guide)
- N2114 ('Understanding NICE guidance').

NICE clinical guidelines are recommendations about the treatment and care of people with specific diseases and conditions in the NHS in England and Wales.

This guidance represents the view of NICE, which was arrived at after careful consideration of the evidence available. Healthcare professionals are expected to take it fully into account when exercising their clinical judgement. However, the guidance does not override the individual responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and/or guardian or carer, and informed by the summary of product characteristics of any drugs they are considering.

Implementation of this guidance is the responsibility of local commissioners and/or providers. Commissioners and providers are reminded that it is their responsibility to implement the guidance, in their local context, in light of their duties to avoid unlawful discrimination and to have regard to promoting equality of opportunity. Nothing in this guidance should be interpreted in a way that would be inconsistent with compliance with those duties.

National Institute for Health and Clinical Excellence

MidCity Place
71 High Holborn
London WC1V 6NA

www.nice.org.uk

© National Institute for Health and Clinical Excellence, 2010. All rights reserved. This material may be freely reproduced for educational and not-for-profit purposes. No reproduction by or for commercial organisations, or for commercial purposes, is allowed without the express written permission of NICE.

Contents

Introduction 4
Person-centred care 6
Key priorities for implementation 7
1 Guidance 10
1.1 Providing information for people with chest pain 10
1.2 People presenting with acute chest pain 11
1.3 People presenting with stable chest pain 20
2 Notes on the scope of the guidance 31
3 Implementation 32
4 Research recommendations 32
4.1 Cost-effectiveness of multislice CT coronary angiography for ruling out obstructive CAD in people with troponin-negative acute coronary syndromes 33
4.2 Novel cardiac biomarkers in people with acute chest pain 34
4.3 Refining the use of telephone advice in people with chest pain 34
4.4 Establishing a national registry for people who are undergoing initial assessment for stable angina 35
4.5 Cost-effectiveness of multislice CT coronary angiography compared with functional testing in the diagnosis of angina 36
4.6 Information about presenting and explaining tests 37
5 Other versions of this guideline 38
5.1 Full guideline 38
5.2 Quick reference guide 38
5.3 'Understanding NICE guidance' 39
6 Related NICE guidance 39
7 Updating the guideline 40
Appendix A: The Guideline Development Group and NICE project team 41
Appendix B: The Guideline Review Panel 45
Appendix C: The algorithms 46

This guidance partially updates NICE technology appraisal guidance 73 (published November 2003).

Recommendation 1.3.6.1 in this guideline replaces recommendation 1.1 of NICE technology appraisal guidance 73. The NICE technology appraisal guidance and supporting documents are available from www.nice.org.uk/guidance/TA73

Introduction

Conditions causing chest pain or discomfort, such as an acute coronary syndrome or angina, have a potentially poor prognosis, emphasising the importance of prompt and accurate diagnosis. Treatments are available to improve symptoms and prolong life, hence the need for this guideline.

This guideline covers the assessment and diagnosis of people with recent onset chest pain or discomfort of suspected cardiac origin. In deciding whether chest pain may be cardiac and therefore whether this guideline is relevant, a number of factors should be taken into account. These include the person's history of chest pain, their cardiovascular risk factors, history of ischaemic heart disease and any previous treatment, and previous investigations for chest pain.

For pain that is suspected to be cardiac, there are two separate diagnostic pathways presented in the guideline. The first is for people with acute chest pain and a suspected acute coronary syndrome, and the second is for people with intermittent stable chest pain in whom stable angina is suspected. The guideline includes how to determine whether myocardial ischaemia is the cause of the chest pain and how to manage the chest pain while people are being assessed and investigated.

As far as possible, the recommendations in this guideline have been listed in the order in which they will be carried out and follow the diagnostic pathways. But, as there are many permutations at each decision point, it has been

necessary to include frequent cross-referencing to avoid repeating recommendations several times.

The algorithms presented in appendix C show the two diagnostic pathways.

This guideline does not cover the diagnosis and management of chest pain that is unrelated to the heart (for example, traumatic chest wall injury, herpes zoster infection) when myocardial ischaemia has been excluded. The guideline also recognises that in people with a prior diagnosis of coronary artery disease, chest pain or discomfort is not necessarily cardiac.

The term ‘chest pain’ is used throughout the guideline to mean chest pain or discomfort.

The guideline will assume that prescribers will use a drug’s summary of product characteristics to inform decisions made with individual patients.

Person-centred care

This guideline offers best practice advice on the care of people who present with recent chest pain or discomfort of suspected cardiac origin.

Treatment and care should take into account people's needs and preferences. People with recent chest pain or discomfort of suspected cardiac origin should have the opportunity to make informed decisions about their care and treatment, in partnership with their healthcare professionals. If people do not have the capacity to make decisions, healthcare professionals should follow the Department of Health's advice on consent (available from www.dh.gov.uk/consent) and the code of practice that accompanies the Mental Capacity Act (summary available from www.publicguardian.gov.uk). In Wales, healthcare professionals should follow advice on consent from the Welsh Assembly Government (available from www.wales.nhs.uk/consent).

Good communication between healthcare professionals and the person with chest pain is essential. It should be supported by evidence-based written information tailored to the person's needs. It should be recognised that the person may be anxious, particularly when the cause of the chest pain is unknown. The options and consequences at every stage of the investigative process should be clearly explained. Investigations, treatment and care, and the information people are given about them, should be culturally appropriate. It should also be accessible to people with additional needs such as physical, sensory or learning disabilities, and to people who do not speak or read English.

If the person agrees, families and carers should have the opportunity to be involved in decisions about treatment and care.

Families and carers should also be given the information and support they need.

Key priorities for implementation

Presentation with acute chest pain

- Take a resting 12-lead electrocardiogram (ECG) as soon as possible. When people are referred, send the results to hospital before they arrive if possible. Recording and sending the ECG should not delay transfer to hospital. **[1.2.2.1]**
- Do not exclude an acute coronary syndrome (ACS) when people have a normal resting 12-lead ECG. **[1.2.2.5]**
- Do not routinely administer oxygen, but monitor oxygen saturation using pulse oximetry as soon as possible, ideally before hospital admission. Only offer supplemental oxygen to:
    - people with oxygen saturation (SpO_2) of less than 94% who are not at risk of hypercapnic respiratory failure, aiming for SpO_2 of 94–98%
    - people with chronic obstructive pulmonary disease who are at risk of hypercapnic respiratory failure, to achieve a target SpO_2 of 88–92% until blood gas analysis is available. **[1.2.3.3]**

- Do not assess symptoms of an ACS differently in ethnic groups. There are no major differences in symptoms of an ACS among different ethnic groups. **[1.2.1.6]**

Presentation with stable chest pain

- Diagnose stable angina based on one of the following:
    - clinical assessment alone **or**
    - clinical assessment plus diagnostic testing (that is, anatomical testing for obstructive coronary artery disease [CAD] and/or functional testing for myocardial ischaemia). **[1.3.1.1]**

- If people have features of typical angina based on clinical assessment and their estimated likelihood of CAD is greater than 90% (see table 1), further diagnostic investigation is unnecessary. Manage as angina. **[1.3.3.5]**

Table 1 Percentage of people estimated to have coronary artery disease according to typicality of symptoms, age, sex and risk factors

	Non-anginal chest pain				Atypical angina				Typical angina			
	Men		Women		Men		Women		Men		Women	
Age (years)	Lo	Hi	Lo	Hi	Lo	Hi	Lo	Hi	Lo	Hi	Lo	Hi
35	3	35	1	19	8	59	2	39	30	88	10	78
45	9	47	2	22	21	70	5	43	51	92	20	79
55	23	59	4	25	45	79	10	47	80	95	38	82
65	49	69	9	29	71	86	20	51	93	97	56	84

For men older than 70 with atypical or typical symptoms, assume an estimate > 90%.

For women older than 70, assume an estimate of 61–90% EXCEPT women at high risk AND with typical symptoms where a risk of > 90% should be assumed.

Values are per cent of people at each mid-decade age with significant coronary artery disease (CAD)[1].

Hi = High risk = diabetes, smoking and hyperlipidaemia (total cholesterol > 6.47 mmol/litre).

Lo = Low risk = none of these three.

The shaded area represents people with symptoms of non-anginal chest pain, who would not be investigated for stable angina routinely.

Note:

These results are likely to overestimate CAD in primary care populations.

If there are resting ECG ST-T changes or Q waves, the likelihood of CAD is higher in each cell of the table.

- Unless clinical suspicion is raised based on other aspects of the history and risk factors, exclude a diagnosis of stable angina if the pain is non-anginal (see recommendation 1.3.3.1). Other features which make a diagnosis of stable angina unlikely are when the chest pain is:
    - continuous or very prolonged **and/or**
    - unrelated to activity **and/or**
    - brought on by breathing in **and/or**
    - associated with symptoms such as dizziness, palpitations, tingling or difficulty swallowing.

  Consider causes of chest pain other than angina (such as gastrointestinal or musculoskeletal pain). **[1.3.3.6]**

[1] Adapted from Pryor DB, Shaw L, McCants CB et al. (1993) Value of the history and physical in identifying patients at increased risk for coronary artery disease. Annals of Internal Medicine 118(2): 81–90.

- In people without confirmed CAD, in whom stable angina cannot be diagnosed or excluded based on clinical assessment alone, estimate the likelihood of CAD (see table 1). Take the clinical assessment and the resting 12-lead ECG into account when making the estimate. Arrange further diagnostic testing as follows:
    - If the estimated likelihood of CAD is 61–90%, offer invasive coronary angiography as the first-line diagnostic investigation if appropriate (see recommendations 1.3.4.4 and 1.3.4.5).
    - If the estimated likelihood of CAD is 30–60%, offer functional imaging as the first-line diagnostic investigation (see recommendation 1.3.4.6).
    - If the estimated likelihood of CAD is 10–29%, offer CT calcium scoring as the first-line diagnostic investigation (see recommendation 1.3.4.7). **[1.3.3.16]**

- Do not use exercise ECG to diagnose or exclude stable angina for people without known CAD. **[1.3.6.5]**

1 Guidance

The following guidance is based on the best available evidence. The full guideline (www.nice.org.uk/guidance/CG95) gives details of the methods and the evidence used to develop the guidance.

1.1 Providing information for people with chest pain

1.1.1.1 Discuss any concerns people (and where appropriate their family or carer/advocate) may have, including anxiety when the cause of the chest pain is unknown. Correct any misinformation.

1.1.1.2 Offer people a clear explanation of the possible causes of their symptoms and the uncertainties.

1.1.1.3 Clearly explain the options to people at every stage of investigation. Make joint decisions with them and take account of their preferences:

- Encourage people to ask questions.
- Provide repeated opportunities for discussion.
- Explain test results and the need for any further investigations.

1.1.1.4 Provide information about any proposed investigations using everyday, jargon-free language. Include:

- their purpose, benefits and any limitations of their diagnostic accuracy
- duration
- level of discomfort and invasiveness
- risk of adverse events.

1.1.1.5 Offer information about the risks of diagnostic testing, including any radiation exposure.

1.1.1.6 Address any physical or learning difficulties, sight or hearing problems and difficulties with speaking or reading English, which may affect people's understanding of the information offered.

1.1.1.7 Offer information after diagnosis as recommended in the relevant disease management guidelines[2].

1.1.1.8 Explain if the chest pain is non-cardiac and refer people for further investigation if appropriate.

1.1.1.9 Provide individual advice to people about seeking medical help if they have further chest pain.

1.2 People presenting with acute chest pain

This section of the guideline covers the assessment and diagnosis of people with recent acute chest pain or discomfort, suspected to be caused by an acute coronary syndrome (ACS). The term ACS covers a range of conditions including unstable angina, ST-segment-elevation myocardial infarction (STEMI) and non-ST-segment-elevation myocardial infarction (NSTEMI).

The guideline addresses assessment and diagnosis irrespective of setting, because people present in different ways. Please note that 'Unstable angina and NSTEMI' (NICE clinical guideline 94) covers the early management of these conditions once a firm diagnosis has been made and before discharge from hospital.

1.2.1 Initial assessment and referral to hospital

1.2.1.1 Check immediately whether people currently have chest pain. If they are pain free, check when their last episode of pain was, particularly if they have had pain in the last 12 hours.

1.2.1.2 Determine whether the chest pain may be cardiac and therefore whether this guideline is relevant, by considering:

- the history of the chest pain
- the presence of cardiovascular risk factors
- history of ischaemic heart disease and any previous treatment
- previous investigations for chest pain.

[2] For example, 'Unstable angina and NSTEMI' (NICE clinical guideline 94), 'Anxiety' (NICE clinical guideline 22) and 'Dyspepsia' (NICE clinical guideline 17).

1.2.1.3 Initially assess people for any of the following symptoms, which may indicate an ACS:

- pain in the chest and/or other areas (for example, the arms, back or jaw) lasting longer than 15 minutes
- chest pain associated with nausea and vomiting, marked sweating, breathlessness, or particularly a combination of these
- chest pain associated with haemodynamic instability
- new onset chest pain, or abrupt deterioration in previously stable angina, with recurrent chest pain occurring frequently and with little or no exertion, and with episodes often lasting longer than 15 minutes.

1.2.1.4 Do not use people's response to glyceryl trinitrate (GTN) to make a diagnosis.

1.2.1.5 Do not assess symptoms of an ACS differently in men and women. Not all people with an ACS present with central chest pain as the predominant feature.

1.2.1.6 Do not assess symptoms of an ACS differently in ethnic groups. There are no major differences in symptoms of an ACS among different ethnic groups.

1.2.1.7 Refer people to hospital as an emergency if an ACS is suspected (see recommendation 1.2.1.3) and:

- they currently have chest pain **or**
- they are currently pain free, but had chest pain in the last 12 hours, and a resting 12-lead ECG is abnormal or not available.

1.2.1.8 If an ACS is suspected (see recommendation 1.2.1.3) and there are no reasons for emergency referral, refer people for urgent same-day assessment if:

- they had chest pain in the last 12 hours, but are now pain free with a normal resting 12-lead ECG **or**
- the last episode of pain was 12–72 hours ago.

1.2.1.9 Refer people for assessment in hospital if an ACS is suspected (see recommendation 1.2.1.3) and:

- the pain has resolved **and**
- there are signs of complications such as pulmonary oedema.

Use clinical judgement to decide whether referral should be as an emergency or urgent same-day assessment.

1.2.1.10 If a recent ACS is suspected in people whose last episode of chest pain was more than 72 hours ago and who have no complications such as pulmonary oedema:

- carry out a detailed clinical assessment (see recommendations 1.2.4.2 and 1.2.4.3)
- confirm the diagnosis by resting 12-lead ECG and blood troponin level
- take into account the length of time since the suspected ACS when interpreting the troponin level.

Use clinical judgement to decide whether referral is necessary and how urgent this should be.

1.2.1.11 Refer people to hospital as an emergency if they have a recent (confirmed or suspected) ACS and develop further chest pain.

1.2.1.12 When an ACS is suspected, start management immediately in the order appropriate to the circumstances (see section 1.2.3) and take a resting 12-lead ECG (see section 1.2.2). Take the ECG as soon as possible, but do not delay transfer to hospital.

1.2.1.13 If an ACS is not suspected, consider other causes of the chest pain, some of which may be life-threatening (see recommendations 1.2.6.5, 1.2.6.6 and 1.2.6.7).

1.2.2 Resting 12-lead ECG

1.2.2.1 Take a resting 12-lead ECG as soon as possible. When people are referred, send the results to hospital before they arrive if possible. Recording and sending the ECG should not delay transfer to hospital.

1.2.2.2 Follow local protocols for people with a resting 12-lead ECG showing regional ST-segment elevation or presumed new left bundle branch block (LBBB) consistent with an acute STEMI until a firm diagnosis is made. Continue to monitor (see recommendation 1.2.3.4).

1.2.2.3 Follow 'Unstable angina and NSTEMI' (NICE clinical guideline 94) for people with a resting 12-lead ECG showing regional ST-segment depression or deep T wave inversion suggestive of a NSTEMI or unstable angina until a firm diagnosis is made. Continue to monitor (see recommendation 1.2.3.4).

1.2.2.4 Even in the absence of ST-segment changes, have an increased suspicion of an ACS if there are other changes in the resting 12-lead ECG, specifically Q waves and T wave changes. Consider following 'Unstable angina and NSTEMI' (NICE clinical guideline 94) if these conditions are likely. Continue to monitor (see recommendation 1.2.3.4).

1.2.2.5 Do not exclude an ACS when people have a normal resting 12-lead ECG.

1.2.2.6 If a diagnosis of ACS is in doubt, consider:

- taking serial resting 12-lead ECGs
- reviewing previous resting 12-lead ECGs

- recording additional ECG leads.

Use clinical judgement to decide how often this should be done. Note that the results may not be conclusive.

1.2.2.7 Obtain a review of resting 12-lead ECGs by a healthcare professional qualified to interpret them as well as taking into account automated interpretation.

1.2.2.8 If clinical assessment (as described in recommendation 1.2.1.10) and a resting 12-lead ECG make a diagnosis of ACS less likely, consider other acute conditions. First consider those that are life-threatening such as pulmonary embolism, aortic dissection or pneumonia. Continue to monitor (see recommendation 1.2.3.4).

1.2.3 Immediate management of a suspected acute coronary syndrome

Management of ACS should start as soon as it is suspected, but should not delay transfer to hospital. The recommendations in this section should be carried out in the order appropriate to the circumstances.

1.2.3.1 Offer pain relief as soon as possible. This may be achieved with GTN (sublingual or buccal), but offer intravenous opioids such as morphine, particularly if an acute myocardial infarction (MI) is suspected.

1.2.3.2 Offer people a single loading dose of 300 mg aspirin as soon as possible unless there is clear evidence that they are allergic to it.

If aspirin is given before arrival at hospital, send a written record that it has been given with the person.

Only offer other antiplatelet agents in hospital. Follow appropriate guidance (‘Unstable angina and NSTEMI’ [NICE clinical guideline 94] or local protocols for STEMI).

1.2.3.3 Do not routinely administer oxygen, but monitor oxygen saturation using pulse oximetry as soon as possible, ideally before hospital admission. Only offer supplemental oxygen to:

- people with oxygen saturation (SpO_2) of less than 94% who are not at risk of hypercapnic respiratory failure, aiming for SpO_2 of 94–98%
- people with chronic obstructive pulmonary disease who are at risk of hypercapnic respiratory failure, to achieve a target SpO_2 of 88–92% until blood gas analysis is available.

1.2.3.4 Monitor people with acute chest pain, using clinical judgement to decide how often this should be done, until a firm diagnosis is made. This should include:

- exacerbations of pain and/or other symptoms
- pulse and blood pressure
- heart rhythm
- oxygen saturation by pulse oximetry
- repeated resting 12-lead ECGs **and**
- checking pain relief is effective.

1.2.3.5 Manage other therapeutic interventions using appropriate guidance ('Unstable angina and NSTEMI' [NICE clinical guideline 94] or local protocols for STEMI).

1.2.4 Assessment in hospital for people with a suspected acute coronary syndrome

1.2.4.1 Take a resting 12-lead ECG and a blood sample for troponin I or T measurement (see section 1.2.5) on arrival in hospital.

1.2.4.2 Carry out a physical examination to determine:

- haemodynamic status
- signs of complications, for example pulmonary oedema, cardiogenic shock **and**

- signs of non-coronary causes of acute chest pain, such as aortic dissection.

1.2.4.3 Take a detailed clinical history unless a STEMI is confirmed from the resting 12-lead ECG (that is, regional ST-segment elevation or presumed new LBBB). Record:

- the characteristics of the pain
- other associated symptoms
- any history of cardiovascular disease
- any cardiovascular risk factors **and**
- details of previous investigations or treatments for similar symptoms of chest pain.

1.2.5 Use of biochemical markers for diagnosis of an acute coronary syndrome

1.2.5.1 Take a blood sample for troponin I or T measurement on initial assessment in hospital. These are the preferred biochemical markers to diagnose acute MI.

1.2.5.2 Take a second blood sample for troponin I or T measurement 10–12 hours after the onset of symptoms.

1.2.5.3 Do not use biochemical markers such as natriuretic peptides and high sensitivity C-reactive protein to diagnose an ACS.

1.2.5.4 Do not use biochemical markers of myocardial ischaemia (such as ischaemia-modified albumin) as opposed to markers of necrosis when assessing people with acute chest pain.

1.2.5.5 Take into account the clinical presentation, the time from onset of symptoms and the resting 12-lead ECG findings when interpreting troponin measurements.

1.2.6 Making a diagnosis

1.2.6.1 When diagnosing MI, use the universal definition of myocardial infarction[3]. This is the detection of rise and/or fall of cardiac biomarkers (preferably troponin) with at least one value above the 99th percentile of the upper reference limit, together with evidence of myocardial ischaemia with at least one of the following:

- symptoms of ischaemia
- ECG changes indicative of new ischaemia (new ST-T changes or new LBBB)
- development of pathological Q wave changes in the ECG
- imaging evidence of new loss of viable myocardium or new regional wall motion abnormality[4].

The clinical classification of MI includes:

- Type 1: spontaneous MI related to ischaemia due to a primary coronary event such as plaque erosion and/or rupture, fissuring or dissection.
- Type 2: MI secondary to ischaemia due to either increased oxygen demand or decreased supply, such as coronary spasm, coronary embolism, anaemia, arrhythmias, hypertension, or hypotension.

1.2.6.2 When a raised troponin level is detected in people with a suspected ACS, reassess to exclude other causes for raised troponin (for example, myocarditis, aortic dissection or pulmonary embolism) before confirming the diagnosis of ACS.

---

[3] Thygesen K, Alpert JS, White HD et al. on behalf of the joint ESC/ACCF/AHA/WHF Task Force for the redefinition of myocardial infarction (2007). Universal definition of myocardial infarction. Journal of the American College of Cardiology 50: 2173–95.

[4] The Guideline Development Group did not review the evidence for the use of imaging evidence of new loss of viable myocardium or new regional wall motion abnormality in the diagnosis of MI, but recognised that it was included as a criterion in the universal definition of MI. The Guideline Development Group recognised that it could be used, but would not be done routinely when there were symptoms of ischaemia and ECG changes.

1.2.6.3 When a raised troponin level is detected in people with a suspected ACS, follow the appropriate guidance (‘Unstable angina and NSTEMI’ [NICE clinical guideline 94] or local protocols for STEMI) until a firm diagnosis is made. Continue to monitor (see recommendation 1.2.3.4).

1.2.6.4 When a diagnosis of ACS is confirmed, follow the appropriate guidance (‘Unstable angina and NSTEMI’ [NICE clinical guideline 94] or local protocols for STEMI).

1.2.6.5 Reassess people with chest pain without raised troponin levels (determined from appropriately timed samples) and no acute resting 12-lead ECG changes to determine whether their chest pain is likely to be cardiac.

If myocardial ischaemia is suspected, follow the recommendations on stable chest pain in this guideline (see section 1.3). Use clinical judgement to decide on the timing of any further diagnostic investigations.

1.2.6.6 Consider a chest X-ray to help exclude complications of ACS such as pulmonary oedema, or other diagnoses such as pneumothorax or pneumonia.

1.2.6.7 Only consider early chest computed tomography (CT) to rule out other diagnoses such as pulmonary embolism or aortic dissection, not to diagnose ACS.

1.2.6.8 If an ACS has been excluded at any point in the care pathway, but people have risk factors for cardiovascular disease, follow the appropriate guidance, for example ‘Lipid modification’ (NICE clinical guideline 67), ‘Hypertension’ (NICE clinical guideline 34).

1.3 People presenting with stable chest pain

This section of the guideline addresses the assessment and diagnosis of intermittent stable chest pain in people with suspected stable angina.

Angina is usually caused by coronary artery disease (CAD). Making a diagnosis of stable angina caused by CAD in people with chest pain is not always straightforward, and the recommendations aim to guide and support clinical judgement. Clinical assessment alone may be sufficient to confirm or exclude a diagnosis of stable angina, but when there is uncertainty, additional diagnostic testing (functional or anatomical testing) guided by the estimates of likelihood of coronary artery disease in table 1 is required.

1.3.1.1 Diagnose stable angina based on one of the following:

- clinical assessment alone **or**
- clinical assessment plus diagnostic testing (that is, anatomical testing for obstructive CAD and/or functional testing for myocardial ischaemia).

1.3.2 Clinical assessment

1.3.2.1 Take a detailed clinical history documenting:

- the age and sex of the person
- the characteristics of the pain, including its location, radiation, severity, duration and frequency, and factors that provoke and relieve the pain
- any associated symptoms, such as breathlessness
- any history of angina, MI, coronary revascularisation, or other cardiovascular disease **and**
- any cardiovascular risk factors.

1.3.2.2 Carry out a physical examination to:

- identify risk factors for cardiovascular disease
- identify signs of other cardiovascular disease

- identify non-coronary causes of angina (for example, severe aortic stenosis, cardiomyopathy) **and**
- exclude other causes of chest pain.

1.3.3 Making a diagnosis based on clinical assessment

1.3.3.1 Anginal pain is:

- constricting discomfort in the front of the chest, or in the neck, shoulders, jaw, or arms
- precipitated by physical exertion
- relieved by rest or GTN within about 5 minutes.

Use clinical assessment and the typicality of anginal pain features listed below to estimate the likelihood of CAD (see table 1):

- Three of the features above are defined as typical angina.
- Two of the three features above are defined as atypical angina.
- One or none of the features above are defined as non-anginal chest pain.

Table 1 Percentage of people estimated to have coronary artery disease according to typicality of symptoms, age, sex and risk factors

	Non-anginal chest pain				Atypical angina				Typical angina			
	Men		Women		Men		Women		Men		Women	
Age (years)	Lo	Hi	Lo	Hi	Lo	Hi	Lo	Hi	Lo	Hi	Lo	Hi
35	3	35	1	19	8	59	2	39	30	88	10	78
45	9	47	2	22	21	70	5	43	51	92	20	79
55	23	59	4	25	45	79	10	47	80	95	38	82
65	49	69	9	29	71	86	20	51	93	97	56	84

For men older than 70 with atypical or typical symptoms, assume an estimate > 90%.

For women older than 70, assume an estimate of 61–90% EXCEPT women at high risk AND with typical symptoms where a risk of > 90% should be assumed.

Values are per cent of people at each mid-decade age with significant coronary artery disease (CAD)[5].

Hi = High risk = diabetes, smoking and hyperlipidaemia (total cholesterol > 6.47 mmol/litre).

Lo = Low risk = none of these three.

The shaded area represents people with symptoms of non-anginal chest pain, who would not be investigated for stable angina routinely.

Note:

These results are likely to overestimate CAD in primary care populations.

If there are resting ECG ST-T changes or Q waves, the likelihood of CAD is higher in each cell of the table.

1.3.3.2 Do not define typical and atypical features of anginal chest pain and non-anginal chest pain differently in men and women.

1.3.3.3 Do not define typical and atypical features of anginal chest pain and non-anginal chest pain differently in ethnic groups.

1.3.3.4 Take the following factors, which make a diagnosis of stable angina more likely, into account when estimating people's likelihood of angina:

- increasing age
- whether the person is male
- cardiovascular risk factors including:
  - a history of smoking

[5] Adapted from Pryor DB, Shaw L, McCants CB et al. (1993) Value of the history and physical in identifying patients at increased risk for coronary artery disease. Annals of Internal Medicine 118(2): 81–90.

  - diabetes
  - hypertension
  - dyslipidaemia
  - family history of premature CAD
  - other cardiovascular disease
- history of established CAD, for example previous MI, coronary revascularisation.

1.3.3.5 If people have features of typical angina based on clinical assessment and their estimated likelihood of CAD is greater than 90% (see table 1), further diagnostic investigation is unnecessary. Manage as angina.

1.3.3.6 Unless clinical suspicion is raised based on other aspects of the history and risk factors, exclude a diagnosis of stable angina if the pain is non-anginal (see recommendation 1.3.3.1). Other features which make a diagnosis of stable angina unlikely are when the chest pain is:

- continuous or very prolonged **and/or**
- unrelated to activity **and/or**
- brought on by breathing in **and/or**
- associated with symptoms such as dizziness, palpitations, tingling or difficulty swallowing.

Consider causes of chest pain other than angina (such as gastrointestinal or musculoskeletal pain).

1.3.3.7 If the estimated likelihood of CAD is less than 10% (see table 1), first consider causes of chest pain other than angina caused by CAD.

1.3.3.8 Consider investigating other causes of angina, such as hypertrophic cardiomyopathy, in people with typical angina-like chest pain and a low likelihood of CAD (estimated at less than 10%).

1.3.3.9 Arrange blood tests to identify conditions which exacerbate angina, such as anaemia, for all people being investigated for stable angina.

1.3.3.10 Only consider chest X-ray if other diagnoses, such as a lung tumour, are suspected.

1.3.3.11 If a diagnosis of stable angina has been excluded at any point in the care pathway, but people have risk factors for cardiovascular disease, follow the appropriate guidance, for example 'Lipid modification' (NICE clinical guideline 67), 'Hypertension' (NICE clinical guideline 34).

1.3.3.12 For people in whom stable angina cannot be diagnosed or excluded on the basis of the clinical assessment alone, take a resting 12-lead ECG as soon as possible after presentation.

1.3.3.13 Do not rule out a diagnosis of stable angina on the basis of a normal resting 12-lead ECG.

1.3.3.14 A number of changes on a resting 12-lead ECG are consistent with CAD and may indicate ischaemia or previous infarction. These include:

- pathological Q waves in particular
- LBBB
- ST-segment and T wave abnormalities (for example, flattening or inversion).

Note that the results may not be conclusive.

Consider any resting 12-lead ECG changes together with people's clinical history and risk factors.

1.3.3.15 For people with confirmed CAD (for example, previous MI, revascularisation, previous angiography) in whom stable angina

cannot be diagnosed or excluded based on clinical assessment alone, see recommendation 1.3.4.8 about functional testing.

1.3.3.16 In people without confirmed CAD, in whom stable angina cannot be diagnosed or excluded based on clinical assessment alone, estimate the likelihood of CAD (see table 1). Take the clinical assessment and the resting 12-lead ECG into account when making the estimate. Arrange further diagnostic testing as follows:

- If the estimated likelihood of CAD is 61–90%, offer invasive coronary angiography as the first-line diagnostic investigation if appropriate (see recommendations 1.3.4.4 and 1.3.4.5).
- If the estimated likelihood of CAD is 30–60%, offer functional imaging as the first-line diagnostic investigation (see recommendation 1.3.4.6).
- If the estimated likelihood of CAD is 10–29%, offer CT calcium scoring as the first-line diagnostic investigation (see recommendation 1.3.4.7).

1.3.3.17 Consider aspirin only if the person's chest pain is likely to be stable angina, until a diagnosis is made. Do not offer additional aspirin if there is clear evidence that people are already taking aspirin regularly or are allergic to it.

1.3.3.18 Follow local protocols for stable angina[6] while waiting for the results of investigations if symptoms are typical of stable angina.

[6] NICE is developing the clinical guideline 'The management of stable angina' (publication expected July 2011).

1.3.4 Diagnostic testing for people in whom stable angina cannot be diagnosed or excluded by clinical assessment alone

This guideline addresses only the diagnostic value of tests for stable angina. The prognostic value of these tests was not considered.

The Guideline Development Group carefully considered the risk of radiation exposure from diagnostic tests. It discussed that the risk needs to be considered in the context of radiation exposure from everyday life, the substantial intrinsic risk that a person will develop cancer during their lifetime and the potential risk of failing to make an important diagnosis if a particular test is not performed. The commonly accepted estimate of the additional lifetime risk of dying from cancer with 10 millisieverts of radiation is 1 in 2000[7]. The Guideline Development Group emphasised that the recommendations in this guideline are to make a diagnosis of chest pain, not to screen for CAD. Most people diagnosed with non-anginal chest pain after clinical assessment need no further diagnostic testing. However in a very small number of people, there are remaining concerns that the pain could be ischaemic, in which case the risk of undiagnosed angina outweighs the risk of any potential radiation exposure.

1.3.4.1 Include the typicality of anginal pain features and the estimate of CAD likelihood (see recommendation 1.3.3.16) in all requests for diagnostic investigations and in the person's notes.

1.3.4.2 Use clinical judgement and take into account people's preferences and comorbidities when considering diagnostic testing.

1.3.4.3 Take into account people's risk from radiation exposure when considering which diagnostic test to use.

[7] Gerber TC et al. (2009) Ionizing radiation in cardiac imaging: a science advisory from the American Heart Association Committee on Cardiac Imaging of the Council on Clinical Cardiology and Committee on Cardiovascular Imaging and Intervention of the Council on Cardiovascular Radiology and Intervention. Circulation 119(7): 1056–65.

1.3.4.4 For people with chest pain in whom stable angina cannot be diagnosed or excluded by clinical assessment alone and who have an estimated likelihood of CAD of 61–90% (see recommendation 1.3.3.16), offer invasive coronary angiography after clinical assessment and a resting 12-lead ECG if:

- coronary revascularisation is being considered **and**
- invasive coronary angiography is clinically appropriate and acceptable to the person.

1.3.4.5 For people with chest pain in whom stable angina cannot be diagnosed or excluded by clinical assessment alone and who have an estimated likelihood of CAD of 61–90% (see recommendation 1.3.3.16), offer non-invasive functional imaging after clinical assessment and a resting 12-lead ECG if:

- coronary revascularisation is not being considered **or**
- invasive coronary angiography is not clinically appropriate or acceptable to the person.

1.3.4.6 For people with chest pain in whom stable angina cannot be diagnosed or excluded by clinical assessment alone and who have an estimated likelihood of CAD of 30–60% (see recommendation 1.3.3.16), offer non-invasive functional imaging for myocardial ischaemia. See section 1.3.6 for further guidance on non-invasive functional testing.

1.3.4.7 For people with chest pain in whom stable angina cannot be diagnosed or excluded by clinical assessment alone and who have an estimated likelihood of CAD of 10–29% (see recommendation 1.3.3.16) offer CT calcium scoring. If the calcium score is:

- zero, consider other causes of chest pain
- 1–400, offer 64-slice (or above) CT coronary angiography
- greater than 400, offer invasive coronary angiography. If this is not clinically appropriate or acceptable to the person and

revascularisation is not being considered, offer non-invasive functional imaging. See section 1.3.6 for further guidance on non-invasive functional testing.

1.3.4.8 For people with confirmed CAD (for example, previous MI, revascularisation, previous angiography), offer non-invasive functional testing when there is uncertainty about whether chest pain is caused by myocardial ischaemia. See section 1.3.6 for further guidance on non-invasive functional testing. An exercise ECG may be used instead of functional imaging.

1.3.5 Additional diagnostic investigations

1.3.5.1 Offer non-invasive functional imaging (see section 1.3.6) for myocardial ischaemia if invasive coronary angiography or 64-slice (or above) CT coronary angiography has shown CAD of uncertain functional significance.

1.3.5.2 Offer invasive coronary angiography as a second-line investigation when the results of non-invasive functional imaging are inconclusive.

1.3.6 Use of non-invasive functional testing for myocardial ischaemia

1.3.6.1 When offering non-invasive functional imaging for myocardial ischaemia use:

- myocardial perfusion scintigraphy with single photon emission computed tomography (MPS with SPECT) **or**
- stress echocardiography **or**
- first-pass contrast-enhanced magnetic resonance (MR) perfusion **or**
- MR imaging for stress-induced wall motion abnormalities.

Take account of locally available technology and expertise, the person and their preferences, and any contraindications when

deciding on the imaging method. [This recommendation updates and replaces recommendation 1.1 of 'Myocardial perfusion scintigraphy for the diagnosis and management of angina and myocardial infarction' (NICE technology appraisal guidance 73)].

1.3.6.2 Use adenosine, dipyridamole or dobutamine as stress agents for MPS with SPECT and adenosine or dipyridamole for first-pass contrast-enhanced MR perfusion.

1.3.6.3 Use exercise or dobutamine for stress echocardiography or MR imaging for stress-induced wall motion abnormalities.

1.3.6.4 Do not use MR coronary angiography for diagnosing stable angina.

1.3.6.5 Do not use exercise ECG to diagnose or exclude stable angina for people without known CAD.

1.3.7 Making a diagnosis following investigations

1.3.7.1 Confirm a diagnosis of stable angina and follow local guidelines for angina[8] when:

- significant CAD (see box 1) is found during invasive or 64-slice (or above) CT coronary angiography **and/or**
- reversible myocardial ischaemia is found during non-invasive functional imaging.

[8] NICE is developing the clinical guideline 'The management of stable angina' (publication expected July 2011).

Box 1 Definition of significant coronary artery disease

Significant coronary artery disease (CAD) found during invasive coronary angiography is ≥ 70% diameter stenosis of at least one major epicardial artery segment or ≥ 50% diameter stenosis in the left main coronary artery:

- Factors intensifying ischaemia.
  Such factors allow less severe lesions (for example ≥ 50%) to produce angina:
  - Reduced oxygen delivery: anaemia, coronary spasm.
  - Increased oxygen demand: tachycardia, left ventricular hypertrophy.
  - Large mass of ischaemic myocardium: proximally located lesions.
  - Longer lesion length.
- Factors reducing ischaemia.
  Such factors may render severe lesions (≥ 70%) asymptomatic:
  - Well developed collateral supply.
  - Small mass of ischaemic myocardium: distally located lesions, old infarction in the territory of coronary supply.

1.3.7.2 Investigate other causes of chest pain when:

- significant CAD (see box 1) is not found during invasive coronary angiography or 64-slice (or above) CT coronary angiography **and/or**
- reversible myocardial ischaemia is not found during non-invasive functional imaging **or**
- the calcium score is zero.

1.3.7.3 Consider investigating other causes of angina, such as hypertrophic cardiomyopathy or syndrome X, in people with typical angina-like chest pain if investigation excludes flow-limiting disease in the epicardial coronary arteries.

2 Notes on the scope of the guidance

NICE guidelines are developed in accordance with a scope that defines what the guideline will and will not cover. The scope of this guideline is available from www.nice.org.uk/guidance/CG95 – click on 'How this guidance was produced'.

The guideline covers adults who have recent onset chest pain or discomfort of suspected cardiac origin, with or without a prior history and/or diagnosis of cardiovascular disease. It includes those presenting with either acute or stable chest pain.

The guideline addresses assessment and investigation irrespective of setting including:

- assessment at initial presentation
- early, initial pharmacological interventions such as oxygen, antiplatelet therapy and pain relief before a cause is known
- choice and timing of investigations
- education and information provision, in particular involving patients in decisions
- where relevant and where associated with chest pain or discomfort, the special needs of people from different groups are considered.

The guideline does not cover the management, including prognostic investigations, and symptom control once the cause of chest pain or discomfort is known. It does not address non-ischaemic chest pain (for example, traumatic chest injury) or pain which is known to be related to another condition, or when there are no cardiac symptoms.

How this guideline was developed

NICE commissioned the National Collaborating Centre for Acute Conditions (now the National Clinical Guideline Centre for Acute and Chronic Conditions) to develop this guideline. The Centre established a Guideline Development Group (see appendix A), which reviewed the evidence and developed the recommendations. An independent Guideline Review Panel oversaw the development of the guideline (see appendix B).

There is more information about how NICE clinical guidelines are developed on the NICE website (www.nice.org.uk/HowWeWork). A booklet, 'How NICE clinical guidelines are developed: an overview for stakeholders, the public and the NHS' (fourth edition, published 2009), is available from NICE publications (phone 0845 003 7783 or email publications@nice.org.uk and quote reference N1739).

3 Implementation

NICE has developed tools to help organisations implement this guidance (see www.nice.org.uk/guidance/CG95).

4 Research recommendations

The Guideline Development Group has made the following recommendations for research, based on its review of evidence, to improve NICE guidance and patient care in the future. The Guideline Development Group's full set of research recommendations is detailed in the full guideline (see section 5).

4.1 Cost-effectiveness of multislice CT coronary angiography for ruling out obstructive CAD in people with troponin-negative acute coronary syndromes

Research question

Is multislice CT coronary angiography a cost-effective first-line test for ruling out obstructive CAD in people with suspected troponin-negative acute coronary syndromes?

Research recommendation

Investigation of the cost-effectiveness of multislice CT coronary angiography as a first-line test for ruling out obstructive CAD in people with suspected troponin-negative acute coronary syndromes.

Why this is important

Current European Society of Cardiology guidelines state that in troponin-negative ACS, with no ST-segment change on the ECG, 'a stress test is recommended … in patients with significant ischaemia during the stress test, coronary angiography and subsequent revascularisation should be considered'. Yet stress testing has relatively low sensitivity and specificity for diagnosing CAD in this group of people. Therefore a significant proportion of at-risk people are missed while others with normal coronary arteries are subjected to an unnecessary invasive coronary angiogram. Multislice CT coronary angiography is highly sensitive and provides a potentially useful means for early rule-out of CAD in troponin-negative acute coronary disease. We need to know whether it is cost effective compared with exercise ECG as a first test in the diagnostic work up of this group.

4.2 Novel cardiac biomarkers in people with acute chest pain

What is the effectiveness and cost effectiveness of new, high-sensitivity troponin assay methods and other new cardiac biomarkers in low, medium, and high risk people with acute chest pain?

Research recommendation

Evaluation of new, high-sensitivity troponin assay methods in low, medium and high risk groups with acute chest pain.

Evaluation of other putative biomarkers compared with the diagnostic and prognostic performance of the most clinically effective and cost-effective troponin assays.

Why this is important

Newer more sensitive troponin assays may offer advantages over previous assays in terms of diagnostic accuracy. They may allow exclusion of myocardial infarction earlier than the 12 hour time frame currently required. Other proposed biomarkers need to be compared to the best available troponin assays.

4.3 Refining the use of telephone advice in people with chest pain

Research question

In what circumstances should telephone advice be given to people calling with chest pain? Is the appropriateness influenced by age, sex or symptoms?

Research recommendation

To develop a robust system for giving appropriate telephone advice to people with chest pain.

Why this is important

The telephone is a common method of first contact with healthcare services, and produces a near uniform emergency response to chest pain symptoms. Such a response has considerable economic, social and human costs. Research should be conducted to clarify if an emergency response in all circumstances is appropriate, or if there are identifiable factors such as age, sex, or associated symptoms that would allow a modified response and a more appropriate use of resources.

Stable chest pain

4.4 Establishing a national registry for people who are undergoing initial assessment for stable angina

Research question and recommendations

Can a national registry of people presenting with suspected angina be established to allow cohort analysis of treatments, investigations and outcomes in this group? Such a registry would provide a vital resource for a range of important research projects, including:

- development and validation of a new score for assessing the estimated likelihood of disease, addressing outstanding uncertainties in the estimation of the likelihood of CAD based on simple measures made at initial assessment (history, examination, routine bloods, resting 12-lead ECG)
- assessment of the extent to which new circulating biomarkers add additional information to measures made at initial assessment
- provision of a framework for trial recruitment without significant work-up bias allowing evaluation of the diagnostic and prognostic test performance of CT-based, MR, echocardiography, and radionuclide technologies.

Why this is important

A national prospective registry of consecutive people with suspected stable angina before initial diagnostic testing does not currently exist in the UK or in any other country. Establishing such a registry would offer the following methodological strengths: statistical size, representative patients without

work-up bias, contemporary data. This would overcome key problems in much of the existing evidence base.

Accurate assessment of the likelihood of coronary disease is needed to inform the cost-effective choice of investigative technologies such as CT coronary calcium scoring for people with chest pain that may be caused by myocardial ischaemia. The data on which the estimated likelihood of CAD is based date from 1979 in a US population and may not be applicable to contemporary UK populations. There remain continuing uncertainties about the initial assessment of people with suspected stable angina. For example, the possible contributions of simple clinical measures such as body mass index, routine blood markers (for example, haemoglobin) or novel circulating biomarkers to estimates of the likelihood of CAD are not known and require further assessment in the whole population and in predefined subgroups including ethnic minorities.

4.5 Cost-effectiveness of multislice CT coronary angiography compared with functional testing in the diagnosis of angina

Research question

What is the clinical and cost effectiveness of multislice CT coronary angiography compared with functional testing in the diagnosis of angina in a population of people with stable chest pain who have a moderate (30–60%) likelihood of CAD?

Research recommendation

Further research should be undertaken to evaluate the clinical and cost effectiveness of multislice CT coronary angiography compared with functional testing in the diagnosis of angina in a population of people with stable chest pain who have a moderate (30–60%) likelihood of CAD.

Why this is important

Multislice CT coronary angiography has developed rapidly in recent years. Published reviews have shown it to be highly effective in the diagnosis of anatomically significant CAD, and costing data indicate that tests can be run at a relatively low cost. However, questions remain about the ability of multislice CT coronary angiography to accurately identify stenoses of functional significance (that is, those that are sufficient to cause angina) in people with stable chest pain. This is especially true for people with a moderate likelihood of significant CAD.

Cost-effectiveness modelling to date has used the diagnosis of CAD as a short-term outcome, and as such inexpensive anatomical tests like multislice CT coronary angiography fare better than functional testing strategies such as MPS with SPECT, stress perfusion MR imaging and stress echocardiography. Because the diagnosis of angina is the true outcome of interest, health economic modelling is needed to evaluate diagnostic technologies on their ability to diagnose stable angina.

4.6 Information about presenting and explaining tests

Research question

All people presenting with chest pain will need to decide whether to accept the diagnostic and care pathways offered. How should information about the diagnostic pathway and the likely outcomes, risks and benefits, with and without treatment, be most effectively presented to particular groups of people, defined by age, ethnicity and sex?

Research recommendation

To establish the best ways of presenting information about the diagnostic pathway to people with chest pain.

Why this is important

Methods of communication (both the content and delivery) will be guided by current evidence-based best practice. Controlled trials should be conducted based on well-constructed randomised controlled clinical trials comparing the

effects of different methods of communication on the understanding of the person with chest pain. Such studies might consider a number of delivery mechanisms, including advice and discussion with a clinician or a specialist nurse as well as specific information leaflets or visual data.

Any trials should also investigate the feasibility of introducing a suggested guideline protocol to be used with all people presenting with chest pain when faced with options concerning their clinical pathway.

Only by clearly explaining and then discussing the proposed diagnostic and care pathways can the healthcare professional be reasonably certain that informed consent has been obtained and that a patient's moral, ethical and spiritual beliefs, expectations, and any misconceptions about their condition, have been taken into account. Consideration should be given to any communication problems the person may have.

5 Other versions of this guideline

5.1 Full guideline

The full guideline, 'Chest pain of recent onset: assessment and diagnosis of recent onset chest pain or discomfort of suspected cardiac origin' contains details of the methods and evidence used to develop the guideline. It is published by the National Clinical Guideline Centre for Acute and Chronic Conditions, and is available from www.rcplondon.ac.uk and our website (http://guidance.nice.org.uk/CG95/Guidance).

5.2 Quick reference guide

A quick reference guide for healthcare professionals is available from http://guidance.nice.org.uk/CG95/QuickRefGuide/pdf/English

For printed copies, phone NICE publications on 0845 003 7783 or email publications@nice.org.uk (quote reference number N2113).

5.3 'Understanding NICE guidance'

A summary for patients and carers ('Understanding NICE guidance') is available from http://guidance.nice.org.uk/CG95/PublicInfo/pdf/English

For printed copies, phone NICE publications on 0845 003 7783 or email publications@nice.org.uk (quote reference number N2114).

We encourage NHS and voluntary sector organisations to use text from this booklet in their own information about chest pain or discomfort of recent onset.

6 Related NICE guidance

Published

- Unstable angina and NSTEMI. NICE clinical guideline 94 (2010). Available from www.nice.org.uk/guidance/CG94
- Lipid modification. NICE clinical guideline 67 (2008). Available from www.nice.org.uk/guidance/CG67
- MI: secondary prevention. NICE clinical guideline 48 (2007). Available from www.nice.org.uk/guidance/CG48
- Hypertension. NICE clinical guideline 34 (2006). Available from www.nice.org.uk/guidance/CG34
- Statins for the prevention of cardiovascular events. NICE technology appraisal guidance 94 (2006). Available from www.nice.org.uk/guidance/TA94
- Anxiety (amended). NICE clinical guideline 22 (2007). Available from www.nice.org.uk/guidance/CG22
- Dyspepsia (amended). NICE clinical guideline 17 (2005). Available from www.nice.org.uk/guidance/CG17
- Myocardial perfusion scintigraphy for the diagnosis and management of angina and myocardial infarction. NICE technology appraisal guidance 73 (2003). Available from www.nice.org.uk/guidance/TA73

Under development

NICE is developing the following guidance (details available from www.nice.org.uk):

- Stable angina. NICE clinical guideline. Publication expected July 2011.
- Prevention of cardiovascular disease. NICE public health guidance. Publication date to be confirmed.

7 Updating the guideline

NICE clinical guidelines are updated so that recommendations take into account important new information. New evidence is checked 3 years after publication, and healthcare professionals and patients are asked for their views; we use this information to decide whether all or part of a guideline needs updating. If important new evidence is published at other times, we may decide to do a more rapid update of some recommendations.

Appendix A: The Guideline Development Group and NICE project team

Guideline Development Group

Professor Adam Timmis (Chair)
Professor of Clinical Cardiology, Barts and the London, Queen Mary's School of Medicine and Dentistry, London

Dr Jane Skinner (Clinical Adviser)
Consultant Community Cardiologist, Royal Victoria Infirmary, Newcastle Upon Tyne

Dr Philip Adams
Cardiologist Consultant, Royal Victoria Infirmary, Newcastle Upon Tyne

Dr John Ashcroft
General Practitioner, Old Station Surgery, Ilkeston, Derbyshire

Ms Liz Clark
Patient representative

Dr Richard Coulden
Consultant Cardiothoracic Radiologist, Glenfield Hospital, Leicester

Professor Harry Hemingway
Public Health Physician Epidemiologist, University College London Medical School, London

Mrs Cathryn James
Clinical Pathways Adviser/Emergency Care Practitioner, Yorkshire Ambulance Service Headquarters, Wakefield

Ms Heather Jarman
Consultant Nurse in Emergency Care, St George's Healthcare NHS Trust, London

Dr Jason Kendall
Consultant in Emergency Medicine, Frenchay Hospital, Bristol

Mr Peter Lewis
Chief Clinical Physiologist, Prince Charles Hospital, Merthyr Tydfil, Wales

Dr Kiran Patel
Consultant Cardiologist and Honorary Senior Lecturer in Cardiovascular Medicine, Sandwell and West Birmingham NHS Trust and University of Birmingham, West Bromwich, West Midlands

Professor Liam Smeeth
Professor of Clinical Epidemiology, London School of Hygiene and Tropical Medicine, London

Mr John Taylor
Patient representative

Nancy Turnbull
Guideline Lead, National Clinical Guideline Centre for Acute and Chronic Conditions

Dr Angela Cooper
Senior Health Services Research Fellow, National Clinical Guideline Centre for Acute and Chronic Conditions

Katrina Sparrow
Health Services Research Fellow, National Clinical Guideline Centre for Acute and Chronic Conditions

Dr Neill Calvert
Head of Health Economics, National Clinical Guideline Centre for Acute and Chronic Conditions

Laura Sawyer
Health Economist, National Clinical Guideline Centre for Acute and Chronic Conditions

David Hill

Project Manager (until December 2009), National Clinical Guideline Centre for Acute and Chronic Conditions

Marian Cotterell

Information Scientist (until January 2009), National Clinical Guideline Centre for Acute and Chronic Conditions

Co-opted GDG Members

Dr Paul Collinson

Consultant in Chemical Pathology and Head of Vascular Risk Management, St George's Hospital, London

Dr Dorothy Frizelle

Clinical Health Psychologist, Department of Clinical Psychology, University of Hull, Hull

Professor Steve Goodacre

Professor of Emergency Medicine, Medical Care Research Unit, Sheffield

Dr Marcus Hardbord

Consultant Physician and Gastroenterologist, Chelsea and Westminster Hospital, London

Ms Helen Williams

Consultant Pharmacist for Cardiovascular Disease, Southwark Health and Social Care, London

NICE project team

Philip Alderson

Associate Director

Sarah Willett

Guideline Commissioning Manager

Andrew Gyton

Guidelines Coordinator

Nichole Taske

Technical Lead

Appendix B: The Guideline Review Panel

The Guideline Review Panel is an independent panel that oversees the development of the guideline and takes responsibility for monitoring adherence to NICE guideline development processes. In particular, the panel ensures that stakeholder comments have been adequately considered and responded to. The panel includes members from the following perspectives: primary care, secondary care, lay, public health and industry.

Dr Rob Walker (Chair)
General Practitioner, Workington

Dr Mark Hill
Head of Medical Affairs, Novartis Pharmaceuticals Ltd

Mrs Ailsa Donnelly
Lay member

Dr John Harley
Clinical Governance and Prescribing Lead and General Practitioner, North Tees PCT

Mr Robin Beal
Consultant in Accident and Emergency Medicine, Isle of Wight

Mrs Sarah Fishburn
Lay member

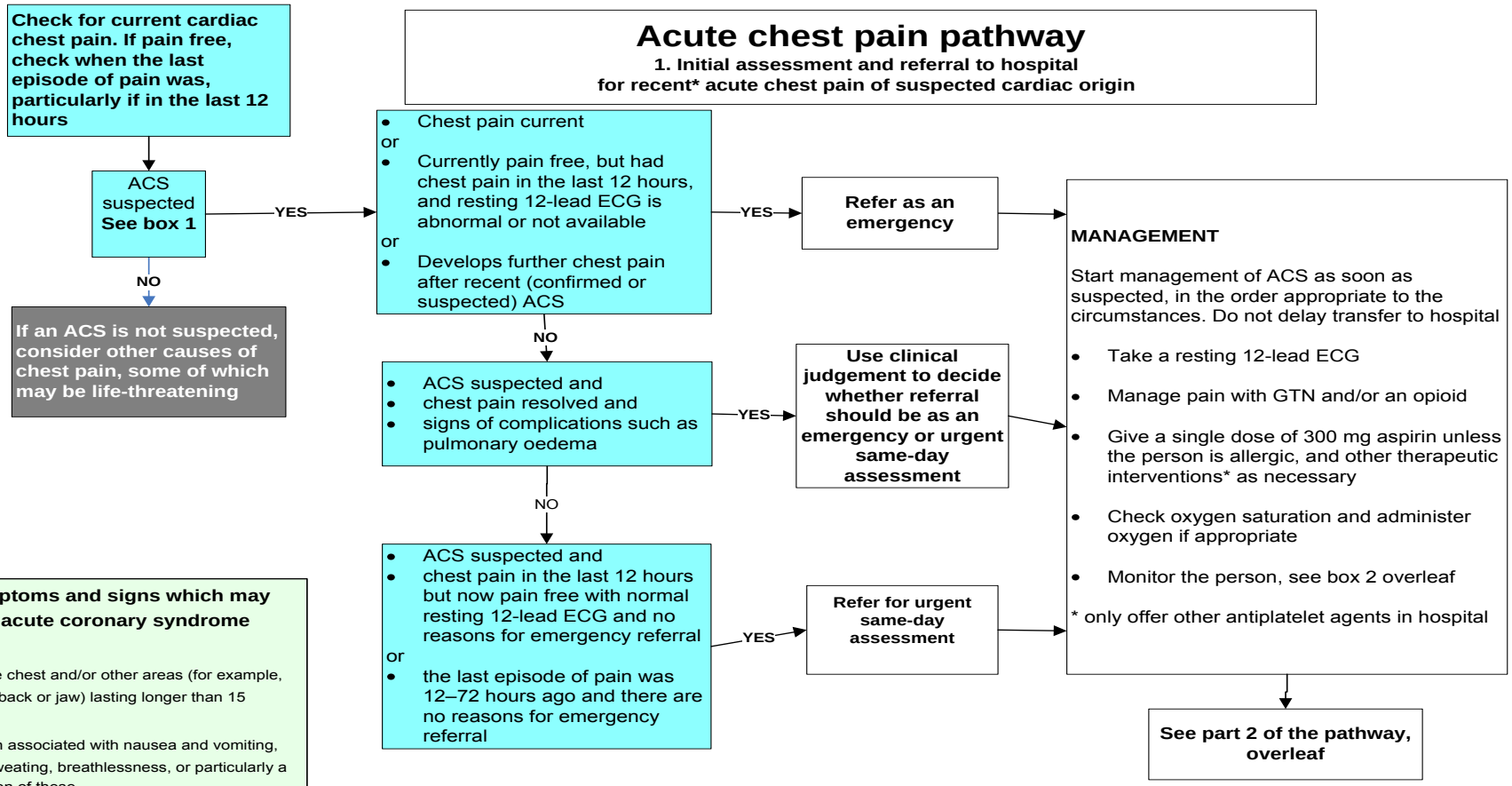
Appendix C: The algorithms

Acute chest pain pathway parts 1 and 2: see pages 47 and 48.

The pathway should be read with the recommendations in this document.

Acute chest pain pathway

1. Initial assessment and referral to hospital for recent* acute chest pain of suspected cardiac origin



Box 1 Symptoms and signs which may indicate an acute coronary syndrome (ACS)

- Pain in the chest and/or other areas (for example, the arms, back or jaw) lasting longer than 15 minutes
- Chest pain associated with nausea and vomiting, marked sweating, breathlessness, or particularly a combination of these
- Chest pain associated with haemodynamic instability
- New onset chest pain, or abrupt deterioration in previously stable angina, with recurrent chest pain occurring frequently and with little or no exertion, and with episodes often lasting longer than 15 minutes

* If a recent ACS is suspected in people whose last episode of chest pain was more than 72 hours ago and who have no complications such as pulmonary oedema: carry out a detailed clinical assessment, confirm the diagnosis by resting 12-lead ECG and blood troponin level (take into account the length of time since the suspected ACS when interpreting the troponin level). Use clinical judgement to decide whether referral is necessary and how urgent this should be

Acute chest pain pathway

2. Investigation and diagnosis in hospital

Assessment in hospital

- Resting 12-lead ECG
- Blood sample for troponin I or T on arrival
- Physical examination
- Clinical history (unless a STEMI is confirmed from the resting 12-lead ECG)

Box 2 Monitoring people with acute chest pain

Use clinical judgement to decide how often this should be done, until a firm diagnosis is made. Include:

- exacerbations of pain and/or other symptoms
- pulse and blood pressure
- heart rhythm
- oxygen saturation by pulse oximetry
- repeated resting 12-lead ECGs and
- checking pain relief is effective.

Box 3 Diagnostic criteria for MI

Rise and /or fall of cardiac biomarkers (preferably troponin) with at least one value above the 99th percentile of the upper reference limit together with evidence of myocardial ischaemia with at least one of the following:

- symptoms of ischaemia
- ECG changes indicative of new ischaemia [new ST-T changes or new left branch bundle block (LBBB)]
- development of pathological Q wave changes in the ECG
- imaging evidence of new loss of viable myocardium or new regional wall motion abnormality.

Resting 12-lead ECG consistent with a STEMI

YES → **Follow local protocols for STEMI until firm diagnosis made (see box 3). Continue to monitor (see box 2)**

NO ↓

Resting 12-lead ECG suggestive of NSTEMI or unstable angina

YES → Follow 'Unstable angina and NSTEMI'* until firm diagnosis made (see box 3). Continue to monitor (see box 2)

NO ↓

- Increase suspicion of an ACS if there are other changes in the resting 12-lead ECG (specifically Q waves, T wave changes)
- Do not exclude an ACS if resting 12-lead ECG is normal
- Consider following 'Unstable angina and NSTEMI'*, if these are very likely. Continue to monitor (see box 2)

If diagnosis of ACS is in doubt:

- Continue monitoring
- Consider taking serial resting 12-lead ECGs, reviewing previous resting 12-lead ECGs and recording additional ECG leads. Use clinical judgement to decide how often this should be done. Note results may not be conclusive
- Repeat troponin measurement 10-12 hours after onset of symptoms
- Consider other acute conditions, firstly life-threatening conditions
- If diagnostic criteria met, follow 'Unstable angina and NSTEMI'* or local protocols for STEMI.

If troponin raised, reassess to exclude other reasons for this → **Follow 'Unstable angina and NSTEMI'* or local protocols for STEMI until firm diagnosis made (see box 3). Continue to monitor (see box 2)**

Take a second blood sample for troponin I or T measurement 10-12 hours after onset of symptoms.

Diagnostic criteria met? See box 3

YES → **Follow 'Unstable angina and NSTEMI'* or local protocols for STEMI**

Uncertain ↓

- Continue to monitor
- Consider chest CT or chest X-ray to exclude other diagnoses/ complications

NO ↓

- **Consider chest CT or chest X-ray to exclude other diagnoses**
- **After reassessment, if myocardial ischaemia is suspected, follow the recommendations on stable chest pain**
- **If an ACS is excluded but people have risk factors for cardiovascular disease, follow the appropriate guidance, for example 'Lipid modification' (NICE clinical guideline 67), 'Hypertension' (NICE clinical guideline 34)**

* NICE clinical guideline 94

Stable chest pain pathway parts 1–3: see pages 50–52.

The pathway should be read with the recommendations in this document.

Stable chest pain pathway

1. Presentation

Carry out a detailed assessment and review

History

Document:

- the age and sex of the person
- the characteristics of the pain and any associated symptoms
- any history of angina, MI, coronary revascularisation, or other cardiovascular disease and
- any cardiovascular risk factors.

Examination

- Identify risk factors and signs of cardiovascular disease
- Identify non-coronary causes of angina (for example, severe aortic stenosis, cardiomyopathy)
- Exclude other causes of chest pain

Person has confirmed CAD

YES

See part 3 of the pathway on page 52

- **Features of pain are non-anginal (see boxes 1 and 2) and**
- **Assessment does not raise clinical suspicion of stable angina**

YES

- **Consider other causes of chest pain**
- **Only consider chest X-ray if other diagnoses are suspected**

NO

Take resting 12-lead ECG (see box 3)

Use clinical assessment and typicality of anginal pain features to stratify the likelihood of CAD (see box 1 and table 1)

Likelihood of CAD is less than 10%

- **Consider other causes of chest pain**
- **Consider investigating other causes of angina such as hypertrophic cardiomyopathy in people with typical angina-like chest pain and a low likelihood of CAD (< 10%)**
- **Only consider chest X-ray if other diagnoses are suspected**

Likelihood of CAD is greater than 90%

- **to identify conditions which exacerbate angina**
- **Treat as stable angina**

Likelihood of CAD is 10- 90%

- **Arrange blood tests to identify conditions which exacerbate angina**
- **Offer further diagnostic testing (see part 2 of pathway on page 51)**
- **Consider aspirin only if the chest pain is likely to be stable angina until diagnosis made**
- **Follow local protocols for stable angina while waiting for the results of investigations if symptoms are typical of stable angina.**

Box 1 Typical stable angina symptoms

- Constricting discomfort in the front of the chest, in the neck, shoulders, jaw, or arms
- Precipitated by physical exertion
- Relieved by rest or GTN within about 5 minutes

Typical angina: all of the above
Atypical angina: two of the above
Non-anginal chest pain: one or none of the above

See recommendation 1.3.3.4 for risk factors which make angina more likely.

Box 2
Stable angina is unlikely if chest pain is:

- continuous or very prolonged and/or
- unrelated to activity and/or
- brought on by breathing in and/or
- associated with symptoms such as dizziness, palpitations, tingling or difficulty swallowing

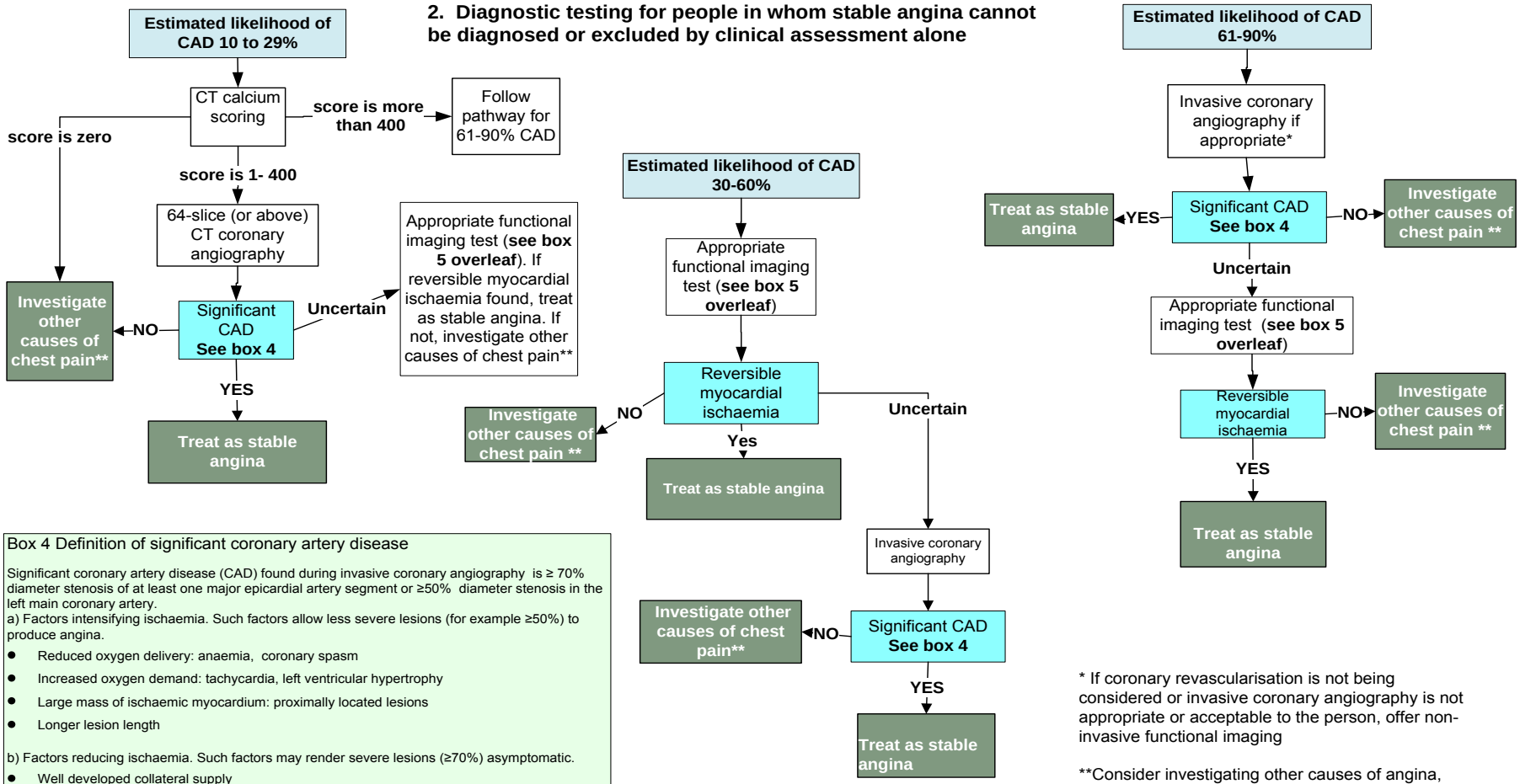
Box 3 Changes on a resting 12-lead ECG consistent with CAD which may indicate ischaemia or previous infarction

- pathological Q waves in particular
- LBBB
- ST-segment and T wave abnormalities (for example, flattening or inversion).

Results may not be conclusive. Consider resting 12-lead ECG changes together with people's clinical history and risk factors. Note that a normal resting 12-lead ECG does not rule out stable angina.

Stable chest pain pathway

2. Diagnostic testing for people in whom stable angina cannot be diagnosed or excluded by clinical assessment alone



Box 4 Definition of significant coronary artery disease

Significant coronary artery disease (CAD) found during invasive coronary angiography is ≥ 70% diameter stenosis of at least one major epicardial artery segment or ≥50% diameter stenosis in the left main coronary artery.

a) Factors intensifying ischaemia. Such factors allow less severe lesions (for example ≥50%) to produce angina.

- Reduced oxygen delivery: anaemia, coronary spasm
- Increased oxygen demand: tachycardia, left ventricular hypertrophy
- Large mass of ischaemic myocardium: proximally located lesions
- Longer lesion length

b) Factors reducing ischaemia. Such factors may render severe lesions (≥70%) asymptomatic.

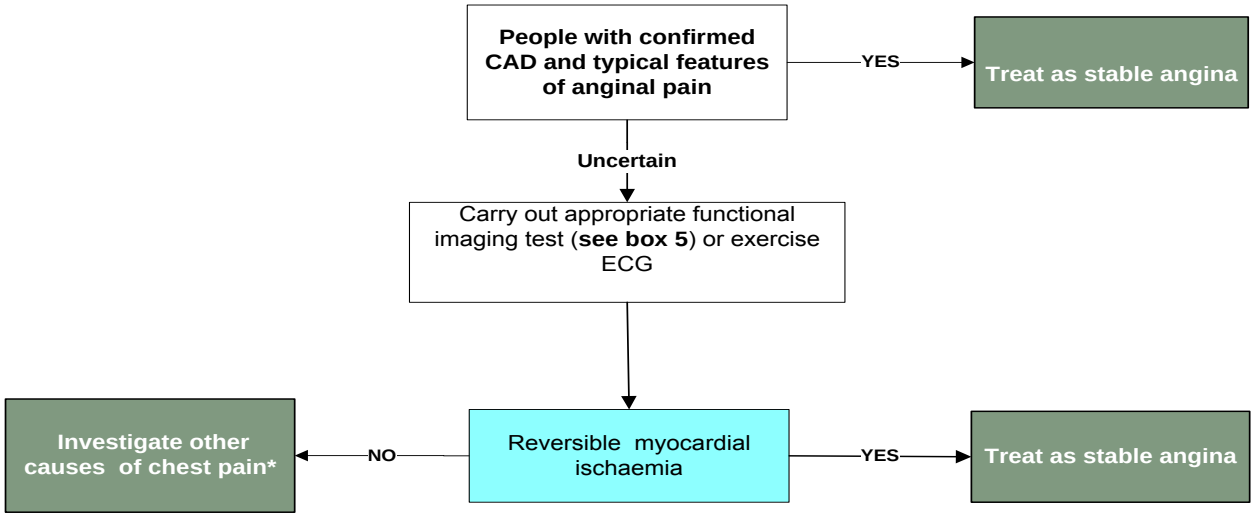
- Well developed collateral supply
- Small mass of ischaemic myocardium: distally located lesions, old infarction in the territory of coronary supply.

* If coronary revascularisation is not being considered or invasive coronary angiography is not appropriate or acceptable to the person, offer non-invasive functional imaging

**Consider investigating other causes of angina, such as hypertrophic cardiomyopathy or syndrome X in people with typical angina-like chest pain if investigation excludes flow-limiting disease in the epicardial coronary arteries.

Stable chest pain pathway

3. Established prior diagnosis of coronary artery disease



Box 5

When offering non-invasive functional imaging for myocardial ischaemia use:

- myocardial perfusion scintigraphy with single photon emission computed tomography (MPS with SPECT) or
- stress echocardiography or
- first-pass contrast-enhanced magnetic resonance (MR) perfusion or
- MR imaging for stress-induced wall motion abnormalities.

Take account of locally available technology and expertise, the person and their preferences, and any contraindications, when deciding on the imaging method.

Note: This recommendation updates and replaces recommendation 1.1 of NICE technology appraisal guidance 73.

* Consider investigating other causes of angina, such as hypertrophic cardiomyopathy or syndrome X in people with typical angina-like chest pain if investigation excludes flow-limiting disease in the epicardial coronary arteries.

HEALTH EVIDENCE REVIEW COMMISSION (HERC)

DRAFT COVERAGE GUIDANCE: CORONARY COMPUTED TOMOGRAPHY ANGIOGRAPHY

DATE: XX/XX/XXXX

HERC COVERAGE GUIDANCE

OPTION1:
Coronary Computed Tomography Angiography (CCTA) should not be covered.

OPTION 2:
CCTA should be covered in the following circumstances:

- In patients with chest pain and an expected Coronary Artery Disease (CAD) prevalence of 50% or less, CCTA should be covered, *if it will lead to either an earlier discharge from the Emergency Department or potentially avoid invasive coronary angiography.*

Patients selected for CCTA: 1) should not be obese; 2) should not have arrhythmias or heart rates more than 65 beats per minute; 3) should be able to hold their breath for more than 20 seconds; 4) should be able to tolerate a standard dose of contrast material; and 5) should not have significant coronary artery calcifications. Multi-slice CT scanners should have at least 64 slices to perform CCTA adequately. The performance and interpretation of CCTA requires special training, and a minimum of 50 cases per year is recommended to maintain competence in the procedure.

CCTA should not be covered for asymptomatic patients.

RATIONALE FOR GUIDANCE DEVELOPMENT

The HERC selects topics for guideline development or technology assessment based on the following principles:

- Represents a significant burden of disease
- Represents important uncertainty with regard to efficacy or harms
- Represents important variation or controversy in clinical care
- Represents high costs, significant economic impact
- Topic is of high public interest

DRAFT

Coverage guidance development follows to translate the evidence review to a policy decision. Coverage guidance may be based on an evidence-based guideline developed by the Evidence-based Guideline Subcommittee or a health technology assessment developed by the Heath Technology Assessment Subcommittee. In addition, coverage guidance may utilize an existing evidence report produced by one of HERC's trusted sources, generally within the last three years.

EVIDENCE SOURCES

Clark, E.E. (2011). *Coronary computed tomographic angiography*. Portland: Center for Evidence-based Policy. Retrieved August 31, 2012, from http://www.ohsu.edu/xd/research/centers-institutes/evidence-based-policy-center/med/index.cfm

National Institute for Health and Clinical Excellence (NICE). (2010*). Chest pain of recent onset: Assessment and diagnosis of recent onset chest pain or discomfort of suspected cardiac origin*. London: NICE. Retrieved August 31, 2012, from http://www.nice.org.uk/nicemedia/live/12947/47938/47938.pdf

Ollendorf, D.A. (2009). *Coronary computed tomographic angiography for the detection of coronary artery disease*. Boston: Institute for Clinical and Economic Review. Retrieved September 18, 2012, from http://www.icer-review.org/index.php/Completed-Appraisals/ccta.html

The summary of evidence in this document is derived directly from these evidence sources, and portions are extracted verbatim.

SUMMARY OF EVIDENCE

Clinical Background

Coronary computed tomographic angiography (CCTA) is a diagnostic imaging test that uses a computed tomographic (CT) scanner to non-invasively image the coronary arteries of the heart. Since obstructive coronary artery disease (CAD) is common in the United States (US) adult population and is responsible for most of the heart attacks, the ability to identify stenosis of the coronary arteries in patients with chest pain becomes important. Coronary computed tomographic angiography can be used in place of other intermediate tests such as stress electrocardiogram (ECG), stress nuclear perfusion imaging and stress echocardiography (ECHO) to either increase or decrease the likelihood of CAD as the cause of chest pain. In contrast to CCTA which provides anatomic information about the coronary arteries, these tests evaluate myocardial ischemia (indicators that the heart muscle is not receiving adequate blood flow).

DRAFT

The development of multi-slice CT scanners has led to increased use of CCTA with nearly half of all cardiology practices in the US leasing or owning cardiac CT equipment. Advocates of CCTA recommend it for patients with low to intermediate risk of CAD who present with acute onset of chest pain (primarily in the emergency department setting) and with stable chest pain suggestive of CAD (primarily in the outpatient setting). Additionally CCTA is being advocated for patients with high risk of CAD and atypical chest pain, evaluation of patients with symptoms after coronary stent placement and screening of asymptomatic patients with high risk of CAD. Both patient selection criteria and equipment capabilities affect the diagnostic efficacy of CCTA. Radiation dose and financial costs for CCTA are significant.

Evidence Review

MED Report (Clark 2011)

Patient and technical factors affect the use and quality of CCTA. Patients selected for CCTA: 1) should not be obese; 2) should not have arrhythmias or heart rates more than 65 beats per minute; 3) should be able to hold their breath for more than 20 seconds; 4) should be able to tolerate a standard dose of contrast material; and 5) should not have significant coronary artery calcifications. Multi-slice CT scanners should have at least 64 slices to perform CCTA adequately. The performance and interpretation of CCTA requires special training, and a minimum of 50 cases per year is recommended to maintain competence in the procedure.

Coronary computed tomographic angiography has a very high sensitivity (≥ 97%) and moderate to moderately high specificity (72-93%) for the detection of coronary artery stenosis, based on moderate quality evidence. A CCTA test sensitivity of 97% means it will detect almost all (97%) of those who have at least one obstructed coronary artery, and only miss 3% of such patients. Thus if the CCTA test is negative it will very likely be a "true negative" and the patients can be sent home. On the other hand, a CCTA test specificity of 72% to 93% means that in a population of patients without obstructive CAD, the test will only be negative 72% to 93% of the time. In the other 7% to 28% of patients without obstructive CAD, it will be a falsely positive test. Practically speaking, a positive CCTA test will often require further testing (invasive angiography) in order to determine if it is a true positive test or a false positive test. These results can be further influenced by the prevalence of obstructive CAD in the population on which the test is used, as described in the body of the report.

These performance characteristics support the use of CCTA to “rule out” obstructive CAD in emergency department patients with acute chest pain and normal ECGs and initial cardiac enzymes, and in outpatients with stable chest pain, a population with low to intermediate probability of obstructive CAD. Coronary computed tomographic

DRAFT

angiography in these situations can be used to identify those patients with no CAD (i.e., negative CCTA in a patient with low to intermediate [pre-test] probability of CAD), so they can be safely discharged from the emergency department without further evaluation. This is substantiated by one small RCT (n = 197) and seven observational studies suggesting that emergency room patients with low to intermediate pre-test probability of CAD and a negative CCTA do not have increased cardiac events over the subsequent year.

In patients with low to intermediate risk of CAD, CCTA appears to have better diagnostic accuracy than stress ECG and stress nuclear perfusion imaging, based on low to moderate quality evidence. A single, poor quality, before and after study suggests that CCTA may reduce the number of subsequent tests including stress nuclear perfusion imaging and invasive coronary angiography. A number of validated clinical prediction rules exist that clinicians can use to assess the [pre-test] probability of obstructive CAD prior to ordering a CCTA.

The MED report did not find studies that addressed screening asymptomatic patients, although they did not specifically search for such evidence.

The amount of radiation dose for CCTA is similar to a CT scan of the abdomen or an invasive coronary angiography, and is estimated to be 8-14 mSv. In addition to radiation exposure and contract reactions or nephropathy, the other potential harms of CCTA are incidental findings. There are relative benefits and harms from the incidental findings noted on CT of the chest (findings in the chest obtained during a CCTA). Approximately 40% to 80% of patients undergoing CCTA will have a finding that is not related to the coronary arteries; 5% to 20% will have a finding deemed clinically important enough for further evaluation. Although some of the patients with these incidental findings will have been judged to have received some benefit, findings from the few studies that have examined this question suggest that the proportion of patients receiving some benefit is very low, while additional risks, anxieties and costs are generated by the additional investigations.

[Evidence Source]

NICE Guideline: Chest Pain of Recent Onset

Acute chest pain (evaluation in the emergency department)
The NICE guideline does not recommend the use of CCTA as a first line test for evaluation of patients in the emergency department with acute chest pain. The guideline assessment of CCTA in this setting is as follows:

In the past few years a number of pilot studies have examined the utility of multislice CT in the emergency department in the differential diagnosis of acute chest pain. To date these studies consist of small numbers of patients (around 100 patients), they have

been conducted primarily in the USA, and they are limited in scope because each represents the experience of one centre. There are differences in study protocols, patient recruitment, scanners used, angiography protocols and angiographic analyses. This makes direct comparison of these studies difficult with respect to reviewing and interpretation. The authors of these studies, while stating the potential promise of multislice CT, do emphasise that further evaluation needs to be done. There are other considerations as given below:

- Currently the use of multislice CCTA in the emergency department would reduce diagnostic time, however this becomes less important with the evolving technology of reduce waiting time for biomarker assay results.
- Multislice CCTA will identify a group of patients with sub clinical CAD i.e. disease that is not the cause of the current chest pain episode. The significance of this will need to be evaluated in large studies in the recruitment of unselected consecutive chest pain patients.
- It has not been established if the patient in the emergency department should receive a dedicated CT coronary angiogram, or have an entire thoracic scan. A dedicated CT coronary angiogram would give the best possible images of the coronary arteries, but allows limited visualisations of other structures that may be responsible for chest pain. The benefit of an entire scan is that it would rule out pulmonary embolism and aortic dissection, however, this would involve increased radiation dose, increased scanning time, and possible less than optimal visualisation of coronary arteries.
- The best use of the multislice CT scanner in the emergency department has not been established. Images could be obtained as soon as possible after initial assessment (history, risk factors, examination) and the first set of cardiac enzymes. In which case the multislice CCTA results would be used as a component of the decision to discharge or admit the patient. Alternatively multislice CCTA could be used to aid in determining what further monitoring and treatment is indicated after a decision has been made to admit the patient. Hence it is unclear at which point multislice CCTA would fit into an algorithm used in the emergency department, and what would be the most cost-effective use of multislice CCTA in the emergency department. This may have implications on cost-effectiveness.
- Current preliminary findings indicate that multislice CCTA in the emergency department has potential for the ruling out of CAD. When stenosis of > 50% is detected the patient would undergo further non invasive or invasive testing, but the precise course of further evaluation is uncertain at this stage due to the limited literature. Resolving this could potentially be a large piece of work, and would impact on the current care pathway.
- Owing to the limited number of studies, health economic evaluation of multislice CCTA in the emergency department may be difficult, particularly as there is no information regarding the subsequent testing of patients when stenosis is > 50%.

Stable Chest Pain (outpatient evaluation)

The NICE guideline makes the following recommendations pertaining to CCTA:

> In people without confirmed CAD, in whom stable angina cannot be diagnosed or excluded based on clinical assessment alone, estimate the likelihood of CAD (see Table 1). Take the clinical assessment and the resting 12-lead ECG into account when making the estimate. Arrange further diagnostic testing as follows:
>
> - If the estimated likelihood of CAD is 61–90%, offer invasive coronary angiography as the first-line diagnostic investigation if appropriate.
> - If the estimated likelihood of CAD is 30–60%, offer functional imaging as the first-line diagnostic investigation.
> - If the estimated likelihood of CAD is 10–29%, offer coronary artery calcium scoring as the first-line diagnostic investigation. If the calcium score is:
> - zero, consider other causes of chest pain
> - 1–400, offer 64-slice (or above) CCTA
> - greater than 400, offer invasive coronary angiography.

Table 1. Percentage of people estimated to have coronary artery disease according to typicality of symptoms, age, sex and risk factors

	Non-anginal chest pain				Atypical angina				Typical angina			
	Men		Women		Men		Women		Men		Women	
Age (years)	**Lo**	**Hi**	**Lo**	**Hi**	**Lo**	**Hi**	**Lo**	**Hi**	**Lo**	**Hi**	**Lo**	**Hi**
35	3	35	1	19	8	59	2	39	30	88	10	78
45	9	47	2	22	21	70	5	43	51	92	20	79
55	23	59	4	25	45	79	10	47	80	95	38	82
65	49	69	9	29	71	86	20	51	93	97	56	84

For men older than 70 with atypical or typical symptoms, assume an estimate > 90%.
For women older than 70, assume an estimate of 61–90% EXCEPT women at high risk AND with typical symptoms where a risk of > 90% should be assumed.

Values are percent of people at each mid-decade age with significant coronary artery disease (CAD)[1].
Hi = High risk = diabetes, smoking and hyperlipidaemia (total cholesterol > 6.47 mmol/litre).
Lo = Low risk = none of these three.
The shaded area represents people with symptoms of non-anginal chest pain, who would not be investigated for stable angina routinely.
Note: These results are likely to overestimate CAD in primary care populations. If there are resting ECG ST-T changes or Q waves, the likelihood of CAD is higher in each cell of the table.

Discussion of the evidence for CCTA in the NICE guideline is summarized as follows:

Advantages and Disadvantages

The advantages of CCTA compared with coronary angiography are that it is less invasive, it can capture thousands of images of a beating heart in seconds, and it may also be relatively less expensive. However 64-slice CCTA requires an injection of

[1] Adapted from Pryor DB, Shaw L, McCants CB et al. (1993) Value of the history and physical in identifying patients at increased risk for coronary artery disease. *Annals of Internal Medicine 118*(2),81-90.

DRAFT

iodine-containing contrast and has been regarded as a moderate to high radiation diagnostic technique (12 to 15 mSv), although recent technical advances are improving radiation efficiency considerably. Further disadvantages of 64-slice CT coronary angiography include; poor correlation with coronary angiography in calcified vessels as extensive calcification obscures imaging of coronary arteries, poor correlation with coronary angiography for quantifying stenosis severity when > 50% and in vessels < 2 mm, no functional assessment of myocardial ischaemia and the potential for motion artifacts due to beating of the heart.

Evidence for Diagnostic Efficacy
For the diagnosis of CAD, five systematic reviews of 64-slice CCTA reported higher sensitivities (ranging from 96% to 99%) and specificities (ranging from 88% to 97%) compared with the non-invasive tests of stress ECHO, stress myocardial perfusion scintigraphy using single photon emission computed tomography (SPECT), stress MR perfusion imaging and stress MR wall motion abnormalities. There is evidence from short term diagnostic economic models that for patients with a low to moderate pre-test likelihood of CAD, 64-slice CCTA (with or without prior exercise ECG) as the initial investigation is cost-effective compared to invasive coronary angiography alone.

Evidence for Risks
The NICE guideline reports on a study that estimated the life attributable risk (LAR) of cancer incidence associated with radiation exposure from 64-slice CCTA. These LARs varied fivefold depending on age and gender, from 1 in 143 for a 20 year old woman to 1 in 3261 for an 80 year old man. The effective dose of radiation from a single scan was reported as a range from 9 to 29 mSv.

Economic Evaluations
Of the six economic evaluations included in evidence reviewed for this guideline, two addressed CCTA. Neither one specified whether they applied to stable or acute chest pain. One compared exercise ECG, dobutamine stress ECHO, dobutamine stress MRI, electron beam CT with calcium scoring and multislice CT coronary angiography as initial diagnostic tests, where only those patients with a positive or indeterminate test result would subsequently undergo invasive coronary angiography (Dewey 2007). Based on this analysis, multislice CT coronary angiography clearly dominates exercise ECG, stress ECHO, stress MRI and calcium scoring with electron beam CT as initial diagnostic strategies for CAD at all levels of disease prevalence modelled. This model did not include any costs for harms of radiation exposure or for evaluation of incidental findings.

The other economic analysis compared 64-slice CCTA compared with exercise ECG, myocardial perfusion scintigraphy with SPECT and invasive coronary angiography in the investigation of CAD (Mowatt 2008). The analysis found that 64-slice CCTA appears to be superior to myocardial perfusion scintigraphy with SPECT for the diagnosis of CAD in all clinical dimensions and also in terms of cost. The report concludes that the high sensitivity and negative predictive value of 64-slice CCTA suggest scope for avoiding unnecessary invasive coronary angiography in those referred for investigation but who do not have CAD. Given the small risk of death associated with invasive coronary angiography, 64-slice CCTA might also confer a small immediate survival

DRAFT

advantage. Avoidance of unnecessary invasive coronary angiography may result in cost savings, even if positive results require confirmation by invasive coronary angiography. However, at higher CAD prevalence, these cost savings are likely to disappear. This model included the costs of complications arising from the interventions, but did not specifically address the harms of radiation or the additional costs of evaluation of incidental findings.

The NICE guideline development group performed their own economic analysis of a diagnostic strategy that incorporated the use of calcium scoring using 64-slice CCTA as a precursor to full 64-slice CCTA. This was done as a way of minimizing the risk of radiation from 64-slice CCTA, a risk which was not explicitly incorporated into the other models. Results of the base case analysis indicate that for lower risk groups (5% and 20%), the use of calcium scoring as a first line testing strategy is likely to be cost-effective and should be followed by either 64-slice CCTA alone or with additional invasive coronary angiography as a confirmatory 3rd test. In higher risk populations, (CAD prevalence greater than 40%), a strategy of sending all patients directly to invasive coronary angiography is likely to be cost-effective. The model indicates that myocardial perfusion scintigraphy with SPECT is excluded through dominance or extended dominance at every level of CAD prevalence. It also indicates that exercise ECG is only cost-effective as a first line investigation strategy at 5% CAD prevalence, but that even in this instance replacing exercise ECG with calcium scoring is likely to improve effectiveness at a reasonable level of additional cost.

[Evidence Source]

Institute for Clinical and Economic Review Report

This cost effectiveness analysis evaluated a variety of diagnostic strategies using stress ECHO, CCTA, SPECT and invasive coronary angiography in two scenarios, in the outpatient setting and in the emergency department assuming either a 30% or 10% prevalence of CAD. All analyses were performed without considering harm, benefit, or costs of radiation-exposure or incidental findings, although they did incorporate an estimate for the evaluation of pulmonary nodules. *"CCTA alone"* resulted in about 14% incidental findings and thus required follow-up as compared to 0% to 5% in the other strategies. Strategies including either CCTA or SPECT as the first or only test exposed all patients to radiation, as opposed to 20% to 40% of patients exposed in strategies with stress-ECHO as the first or only test.

Asymptomatic patients
Use of CCTA as a screening tool in asymptomatic patients was not evaluated in this report.

DRAFT

Emergency department patients with chest pain
When used as triage in the emergency department, they found that at a CAD prevalence of 30%, CCTA produces a higher number of true positives and fewer false negatives relative to other 1- or 2-test strategies, and lower diagnostic phase costs than nearly all other tests; at a prevalence of 10%, differences in test performance are diminished but the pattern of costs remains the same. When alternative estimates of CCTA's diagnostic accuracy are employed, the balance of false-positive and false-negative shifts, but has little impact on comparative cost between the strategies. However, when a more aggressive strategy for management of mild-moderate stenosis is employed, CCTA becomes more costly than several other strategies due to a higher rate of referral for invasive coronary angiography.

In 2012, ICER updated this report to incorporate the findings of two large, multicenter randomized clinical trials of CCTA vs. standard emergency department evaluation. These trials enrolled nearly 2500 patients at 14 sites, and unlike the earlier trial, included patients at intermediate risk of acute coronary syndromes. Findings were very similar between the two studies. Coronary computed tomography angiography was found to significantly increase the percentage of patients discharged home from the emergency department relative to standard care, and reduced time in hospital by seven to eight hours on average. There were no deaths at 28 to 30 days in either study, and no statistically-significant differences in rates of major cardiovascular events. In one study, however, patients in the CCTA arm received more downstream diagnostic testing than those receiving standard evaluation; the increased costs from additional testing eliminated any savings from earlier discharge in the CCTA arm, and average total strategy costs were found to be similar between the groups.

The Institute for Clinical and Economic Review previously found the evidence on comparative clinical effectiveness to be "Comparable" between CCTA and standard triage care in the emergency department setting; these recent findings confirm the original rating. The original rating for comparative value was "High", however, based primarily on evidence of earlier emergency department discharge. In light of these recent data on increased resource use following CCTA, we would recommend changing CCTA's comparative value rating to "Reasonable/Comparable".

Outpatients with chest pain
In the outpatient model, at a CAD prevalence of 30%, CCTA produces a higher number of true positives and fewer false negatives relative to other 1- or 2-test strategies, and lower diagnostic phase costs than nearly all other tests; at a prevalence of 10%, differences in test performance are diminished but the pattern of costs remains the same. When alternative estimates of CCTA's diagnostic accuracy are employed, the balance of false-positive and false-negative shifts, but has little impact on comparative cost between the strategies. However, when a more aggressive strategy for management of mild-moderate stenosis is employed, CCTA becomes more costly than several other strategies due to a higher rate of referral for invasive coronary angiography.

Considering a lifetime horizon, quality-adjusted life expectancy is quite similar across the strategies, with a difference of only about two weeks between the most and least

effective strategies. At 30% CAD prevalence, a single-test strategy with CCTA appears to be more effective and less costly than SPECT, and a reasonable value when compared to Stress ECHO (incremental cost-effectiveness ratios of $13,000 to $16,000/QALY). When prevalence is reduced to 10%, however, while cost-effectiveness is similar for CCTA vs. stress ECHO, SPECT is more effective than CCTA at a ratio of approximately $80,000/QALY. A shift from conservative to aggressive management of mild-moderate stenosis affects the lifetime results only marginally, as does the use of alternative estimates of CCTA's diagnostic accuracy.

Because the range of effectiveness results is so narrow, the model is highly sensitive to changes in selected parameters, in particular the costs of the various strategies. For example, at a cost of $248 or less, CCTA would dominate all other strategies, while for CCTA costs of $1,083, $1,916, and $2,749, the cost-effectiveness ratios would be $50,000/QALY, $100,000/QALY, and $150,000/QALY, respectively.

[Evidence Source]

Overall Summary

Coronary computed tomographic angiography may be useful to "rule out" obstructive CAD in emergency department patients with acute chest pain and normal ECGs and initial cardiac enzymes, and in outpatients with stable chest pain in a population with low to intermediate probability of obstructive CAD. Cost-effectiveness analyses show either that CCTA is comparable or less costly than other diagnostic strategies, although for the most part, they did not consider the economic consequences of the harms of radiation or further evaluation of incidental findings. Use in other patient populations is not recommended due to unacceptable false positive or false negative results. Use in asymptomatic patients has not been evaluated.

PROCEDURE

Coronary computed tomographic angiography

DIAGNOSES

Coronary artery disease
Chest pain

APPLICABLE CODES

CODES	DESCRIPTION
ICD-9 Diagnosis Codes	
410	Acute myocardial infarction
411	Other acute and subacute forms of ischemic heart disease
413	Angina pectoris
414	Other forms of chronic ischemic heart disease
ICD-9 Volume 3 (Procedure Codes)	
87.41	Computed axial tomography of the heart

DRAFT

CPT Codes	
75574	Computed tomographic angiography, heart, coronary arteries and bypass grafts, with contrast, including 3D image post-processing
HCPCS Codes	
None	

Note: Inclusion on this list does not guarantee coverage

Coverage guidance is prepared by the Health Evidence Review Commission (HERC), HERC staff, and subcommittee members. The evidence summary is prepared by the Center for Evidence-based Policy at Oregon Health & Science University (the Center). This document is intended to guide public and private purchasers in Oregon in making informed decisions about health care services.

The Center is not engaged in rendering any clinical, legal, business or other professional advice. The statements in this document do not represent official policy positions of the Center. Researchers involved in preparing this document have no affiliations or financial involvement that conflict with material presented in this document.

DRAFT

Coronary Computed Tomographic Angiography

August 2011

Center for Evidence-based Policy
Medicaid Evidence-based Decisions Project (MED)
Oregon Health & Science University
3455 SW US Veterans Hospital Road
Mailstop SN-4N, Portland, OR 97239-2941
Phone: 503.494.2182
Fax: 503.494.3807
http://www.ohsu.edu/ohsuedu/research/policycenter/med/index.cfm

This report was prepared by:

Edgar E. Clark, MD, MHA

Center for Evidence-based Policy
Oregon Health and Science University

This report was reviewed by:

Martha Gerrity, MD, PhD, MPH

Center for Evidence-based Policy
Oregon Health and Science University

This report was reviewed by experts in the field. Names of reviewers and disposition of review comments are available upon request.

This document was prepared by the Center for Evidence-based Policy at Oregon Health & Science University (the Center). This document is intended to support Medicaid Evidence-based Decisions Project (MED) participant organizations and their constituent decision-making bodies to make informed decisions about the provision of health care services. In addition, this document is available to the public, as approved by the MED governance group. The document is intended as a reference and is provided with the understanding that the Center is not engaged in rendering any clinical, legal, business or other professional advice.

The statements in this document do not represent official policy positions of the Center, the MED Project or MED participating organizations. Researchers and authors involved in preparing this document have no affiliations or financial involvement that conflict with material presented in this document.

About the Center for Evidence-based Policy and the Medicaid Evidence-based Decisions Project

The Center for Evidence-based Policy (http://www.ohsu.edu/policycenter) is a collaboration of academic and governmental entities with the mission of addressing policy challenges with evidence and collaboration. Established in 2003, the Center builds on lessons learned from improving public policy in the field of health care through innovation, collaboration, and use of best evidence.

The Medicaid Evidence-based Decisions (MED) Project is housed at the Center, and is a collaboration of state Medicaid programs and their state partners whose purpose is to make high quality evidence available to states to support benefit design and coverage decisions. The MED Project products include identification and synthesis of high quality systematic reviews and other studies, technology assessments of existing and emerging health technologies, and rapid evaluations of healthcare interventions. Additional information about the project can be found at the following link: http://www.ohsu.edu/xd/research/centers-institutes/evidence-based-policy-center/med/index.cfm.

Suggested citation:
Clark, E.E. (2011). *Coronary Computed Tomographic Angiography*. Portland: Center for Evidence-based Policy. Available at: http://www.ohsu.edu /xd/research/centers-institutes/evidence-based-policy-center/med/index.cfm

Table of Contents

Executive Summary 1

Background 4

Table 1. Prevalence of coronary artery stenosis by character of chest pain 4

Table 2. Conditions that cause chest pain 4

Table 3. Post-test probability of obstructive CAD using the same test but varying the pre-test probability of CAD 5

Table 4. Costs of CCTA and ICA 7

Key Questions 8

Methods 8

Findings 10

Key Question #1: What is the sensitivity and specificity of CT angiography in diagnosing obstructive coronary artery disease compared to catheter-based angiography with or without clinical follow-up? 11

Key Question #2: Are there patients, situations or setting where the results of CT angiography would preclude the use of catheter-based angiography without changing clinical outcomes? .. 13

Key Question #3: What are the rates of revascularization procedures, hospitalizations, coronary heart disease (CHD) events following CT angiography compared to catheter-based angiography? 15

Key Question #4: What is the evidence for harms related to CT angiography compared to catheter-based angiography? 15

Table 5. Relative amounts of radiation exposure from different sources 15

Guidelines 16

Summary 18

Policy Considerations 19

Appendix A. Updated Search Strategy 22

Appendix B. Sensitivity and specificity of CT angiography in diagnosing obstructive coronary artery disease compared to catheter-based angiography with or without clinical follow-up 24

Appendix C. Subgroups of patients, situations or settings where the results of CCTA would preclude the use of catheter-based angiography without changing clinical outcomes 26

Appendix D. Rates of revascularization procedures, hospitalizations and utilization of other diagnostic tests following CT angiography compared to catheter-based angiography 28

Appendix E. Summary of guidelines 29

Appendix F. Number of asymptomatic patients with true positive and false positive results on CT angiography based on prevalence of CAD in a population..... 31

Appendix G. Analysis of Select Medicare, State and Private Payor CCTA Coverage Policies 34

Appendix H. Select Medicare, State and Private Payor CCTA Coverage Policies 35

Appendix I. Relevant codes..... 55

References 56

Abbreviations Used in this Report

ACR. American College of Radiology.

ACS. Acute coronary syndrome = myocardial infarction or unstable angina.

CAD. Obstructive coronary artery disease; stenosis or blockage of the coronary arteries.

CCTA. Coronary computed tomographic angiography. Also referred to in the literature as computed tomographic coronary angiography (CTCA) or coronary CTA.

CPG. Clinical practice guideline.

CT. Computed tomography.

ED. Emergency department or emergency room.

ECG. Electrocardiogram.

ICA. Invasive coronary angiography also catheter coronary angiography.

High probability of CAD. Probability of greater than 80% or 90% (varies by author).

Intermediate probability of CAD. Probability from 10% to 90% or from 20% to 80% (varies by author).

Low probability of CAD. Probability less than 10% or 20% (varies by author).

MA. Meta-analysis.

MACE. Major adverse cardiac event—cardiac death, non-fatal myocardial infarction and coronary revascularization.

MSCT. Multiple slice computed tomography. This refers to modern CT scanner that take more than one slice at a time—in many cases for CCTA the scanners take 64 slices. The modern CT scanners are often referred to as 64 slice CT scanners.

NPI. Nuclear perfusion imaging. A nuclear medicine test to measure the distribution of a radioactive tracer in the myocardium or the regional functioning of the myocardium. This is also referred to as stress NPI and SPECT NPI.

PCI. Percutaneous coronary intervention; this normally refers to placement of a stent in the coronary artery.

RCT. Randomized controlled trial.

Sensitivity. The ability of a test to identify correctly people with a condition. A test with high sensitivity will nearly always be positive for people who have the condition (the test has a low rate of false-negative results). Sensitivity is also known as the true-positive rate.

Specificity. The ability of a test to identify correctly people without a condition. A test with high specificity will rarely be wrong about who does NOT have the condition (the test has a low rate of false-positive results). Specificity is also known as the true-negative rate.

SR. Systematic review.

SPECT. Single photon emission computed tomography; a nuclear medicine test usually used to

perform nuclear perfusion imaging (NPI).

Stress ECG, Stress Echo, Stress SPECT. Stress refers to tests that are performed after a period of exercise on a treadmill.

TA. Technology assessment.

Executive Summary

Background

Coronary computed tomographic angiography (CCTA) is a diagnostic imaging test that uses a computed tomographic (CT) scanner to non-invasively image the coronary arteries of the heart. Since obstructive coronary artery disease (CAD) is common in the United States (US) adult population and is responsible for most of the heart attacks, the ability to identify stenosis of the coronary arteries in patients with chest pain becomes important. Coronary computed tomographic angiography can be used in place of other intermediate tests such as stress electrocardiogram (ECG), stress nuclear perfusion imaging (NPI) and stress echocardiography (ECHO) to either increase or decrease the likelihood of CAD as the cause of chest pain. In contrast to CCTA which provides anatomic information about the coronary arteries, these tests evaluate myocardial ischemia (indicators that the heart muscle is not receiving adequate blood flow). The development of multi-slice CT scanners has led to increased use of CCTA with nearly half of all cardiology practices in the US leasing or owning cardiac CT equipment. Advocates of CCTA recommend it for patients with low to intermediate risk of CAD who present with acute onset of chest pain (primarily in the emergency department setting) and with stable chest pain suggestive of CAD (primarily in the outpatient setting). Additionally CCTA is being advocated for patients with high risk of CAD and atypical chest pain, evaluation of patients with symptoms after coronary stent placement and screening of asymptomatic patients with high risk of CAD.

Both patient selection criteria and equipment capabilities affect the diagnostic efficacy of CCTA. Radiation dose and financial costs for CCTA are significant.

Methods

The key questions addressed in this review include:

> **Key Question #1:** What is the sensitivity and specificity of CT angiography in diagnosing obstructive coronary artery disease compared to catheter-based angiography with or without clinical follow-up?
>
> **Key Question #2:** Are there patients, situations or setting where the results of CT angiography would preclude the use of catheter-based angiography without changing clinical outcomes?
>
> **Key Question #3:** What are the rates of revascularization procedures, hospitalizations and utilization of other diagnostic tests following CT angiography compared to catheter-based angiography?
>
> **Key Question #4:** What is the evidence for harms related to CT angiography compared to catheter-based angiography?

A search of the MED clinical evidence core sources was done to identify systematic reviews (SRs), meta-analyses (MAs), technology assessments (TAs), and clinical practice guidelines (CPGs). A MEDLINE search was done to identify studies published from June 2009, the last search date for a previously identified high quality SR, through February 2010. Our search

located two SRs, one TA that includes a SR, one randomized controlled trial (RCT) and 13 observational studies. Five guidelines were identified.

Findings

Patient and technical factors affect the use and quality of CCTA. Patients selected for CCTA: 1) should not be obese; 2) should not have arrhythmias or heart rates more than 65 beats per minute; 3) should be able to hold their breath for more than 20 seconds; 4) should be able to tolerate a standard dose of contrast material; and 5) should not have significant coronary artery calcifications. Multi-slice CT scanners should have at least 64 slices to perform CCTA adequately. The performance and interpretation of CCTA requires special training, and a minimum of 50 cases per year is recommended to maintain competence in the procedure.

Coronary computed tomographic angiography has a very high sensitivity (≥ 97%) and moderate to moderately high specificity (72-93%) for the detection of coronary artery stenosis, based on moderate quality evidence. A CCTA test sensitivity of 97% means it will detect almost all (97%) of those who have at least one obstructed coronary artery, and only miss 3% of such patients. Thus if the CCTA test is negative it will very likely be a "true negative" and the patients can be sent home. On the other hand, a CCTA test specificity of 72% - 93% means that in a population of patients without obstructive CAD the test will only be negative 72% to 93% of the time. In the other 7% to 28% of patients without obstructive CAD, it will be a falsely positive test. Practically speaking, a positive CCTA test will often require further testing in order to determine if it is a true positive test or a false positive test. These results can be further influenced by the prevalence of obstructive CAD in the population on which the test is used, as described in the body of the report.

These performance characteristics support the use of CCTA to “rule out” obstructive CAD in emergency department (ED) patients with acute chest pain and normal ECGs and initial cardiac enzymes, and in outpatients with stable chest pain, a population with low to intermediate probability of obstructive CAD. Coronary computed tomographic angiography in these situations can be used to identify those patients with no CAD (i.e., negative CCTA in a patient with low to intermediate [pre-test] probability of CAD), so they can be safely discharged from the ED without further evaluation. This is substantiated by one small RCT (n = 197) and 7 observational studies suggesting that *emergency room patients with low to intermediate pre-test probability of CAD and a negative CCTA do not have increased cardiac events over the subsequent year*.

In patients with low to intermediate risk of CAD, CCTA appears to have better diagnostic accuracy than stress ECG and stress NPI, based on low to moderate quality evidence. A single, poor quality, before and after study suggests that CCTA may reduce the number of subsequent tests including stress NPI and ICA. A number of validated clinical prediction rules exist that clinicians can use to assess the [pre-test] probability of obstructive CAD prior to ordering a CCTA.

Although two specialty-based guidelines find CCTA “appropriate” for patients with chest pain and a high risk of CAD, and for screening of high risk asymptomatic patients, other guidelines

recommend against these uses. We found no studies that specifically address these uses of CCTA. Moreover, the pre-test probabilities of CAD in these patient groups might result in a number of false negatives (in symptomatic patients at high risk of CAD) and false positives (in asymptomatic patients at low to intermediate risk of CAD).

Limitations of the Evidence

The overall quality of evidence for the diagnostic accuracy of CCTA is moderate due to moderate risk of bias from the use of convenience samples of patients already scheduled for ICA. Patients scheduled for ICA are likely to have a higher (pre-test) probability of CAD than patients who would receive CCTA in practice. This causes spectrum bias. The estimates of sensitivity and specificity reported in these studies may be higher than would be expected for a more typical patient population undergoing CCTA. The evidence on the effect of CCTA on subsequent cardiac diagnostic testing and cardiac events is from one small fair quality RCT and 13 cohort and case series studies of poor to good quality.

Background

Clinical overview

Epidemiology of coronary artery disease: In the United States (US), heart disease is the leading cause of death for both men and women, with more than one out of four deaths attributable to heart disease (Center for Disease Control and Prevention [CDC] 2010). Coronary artery disease (CAD), which is the most common type of heart disease, was the cause of 445,687 US deaths in 2005 (CDC 2010). It is estimated that 785,000 US adults in 2009 had new myocardial infarctions (heart attacks), with 470,000 US adults having experienced recurrent attacks (Lloyd-Jones 2009). Caucasian and African Americans have the highest percentages of deaths due to heart disease, 27.2% and 25.8% respectively (CDC 2010). Although CAD is the most common form of heart disease, the prevalence of obstructive CAD (stenosis of 50% or more) varies based on age, sex, and symptoms.

Prevalence of obstructive CAD is greater for males than females and increases with age. For example, in asymptomatic individuals, the prevalence ranges from 0.3% for young women to 12.3% for men over 60 years. The prevalence (or pre-test probability) of obstructive CAD also varies depending on the characteristics of the chest pain (Table 1).

Table 1. Prevalence of Coronary Artery Stenosis by Character of Chest Pain*

Character of chest pain	Prevalence of coronary stenosis (± standard error)
Typical angina 1) substernal pressure like chest discomfort 2) provoked by exertion or emotional stress and 3) relieved by rest or nitroglycerin	88.9% (0.7%)
Atypical Angina Chest pain that meets two of the three criteria for "typical" angina (above)	49.9% (1.1%)
Non-anginal chest pain Chest pain that meets one or none of the criteria for "typical" angina (above)	16.0% (1.2%)

* Based on coronary catheterization data reported in Diamond (1979)

The variability in prevalence of CAD based on the character of the chest pain is likely due to the fact that a variety of conditions cause chest pain. Table 2, adapted from Panju (1998), lists common conditions that cause chest pain.

Table 2. Conditions that Cause Chest Pain

Cardiac Causes of Chest Pain	Non-cardiac Causes of Chest Pain
Ischemic	**Gastroesophageal**
Angina	Gastroesophageal reflux (GERD)
Unstable angina	Esophageal spasm
Myocardial infarction	Peptic ulcer disease

Cardiac Causes of Chest Pain	Non-cardiac Causes of Chest Pain
Non-ischemic	**Non-gastroesophageal**
Pericarditis	Pneumothorax
Valvular heart disease	Pulmonary Embolism
Aortic dissection	Musculoskeletal
	Somatoform (panic attack)

Diagnosis of CAD: Because the clinical diagnosis of CAD by history, physical examination and resting ECG alone can be difficult, a number of tests have been developed to aid in earlier and more accurate diagnosis. These tests have included clinical risk scores such as the Duke scores (Bayliss 2002; Gibbons 1999; Pryor 1983, 1993), TIMI risk score (Hess 2010, Than 2011), stress ECG testing, stress nuclear perfusion imaging (NPI) and stress echocardiography (ECHO). The purpose of these tests is to either increase or decrease the likelihood of CAD as the cause of chest pain. Stress ECG, NPI and ECHO are sometimes called functional tests since they attempt to link the symptoms of chest pain with evidence of myocardial ischemia. Coronary computed tomographic angiography (CCTA) is another intermediate test that may be substituted for these tests. However, *CCTA is considered a structural or anatomic test. It can only determine the location and extent of obstruction of a coronary artery; it cannot link chest pain symptoms to the obstructions*.

An important concept in understanding how these tests are employed is the relationship of test sensitivity, specificity, pre-test probability and post-test probability. *The probability of a patient having a condition after completion of a test is related to the prevalence of the condition in the population (and/or pre-test probability based on risk assessment) and to the sensitivity and specificity of the test itself.* Table 3 gives results for four hypothetical patients coming from populations with different prevalence of CAD (5%, 16%, 50% and 89%) and a test with a sensitivity of 70% and specificity of 90%, based on the example in Diamond (1997).

Table 3. Post-test Probability of Obstructive CAD Using the Same Test but Varying the Pre-test Probability of CAD

Pre-test probability (prevalence of disease in the population)	Sensitivity*	Specificity*	Post-test probability if test "positive"	Post-test probability if test "negative"
5%	70%	90%	25%	2%
16%	70%	90%	57%	6%
50%	70%	90%	88%	25%
89%	70%	90%	98%	73%

* These values are *not* for CCTA. Sensitivity and specificity of CCTA are described in the findings.

This concept is important because the *pre-test probability and/or prevalence of a condition can have an important impact on the probability of disease after the test result is known (post-test probability).* Many people presenting with chest pain have a low pre-test probability of CAD (e.g., 5%); *a positive test result increases the probability of CAD* (25% in the example) and may

justify additional testing. A negative test result in patients with a low pre-test probability decreases the probability of CAD to 2%, effectively "ruling out" CAD and allowing discharge of the patient, or pursuit of another diagnosis. In contrast, for a person presenting with chest pain and an intermediate pre-test probability of CAD (e.g., 50%), a negative test will have a post-probability of CAD of 25%. Although a negative test result lowered the probability of CAD from 50% to 25%, a 25% probability may still be high enough to warrant further testing.

Invasive coronary angiography (ICA): Invasive coronary angiography (ICA) is a procedure used to visualize the coronary arteries. The process involves inserting a catheter into a blood vessel in the upper leg or arm and positioning the tip of the catheter at the orifice of a coronary artery. A contrast medium that is visible on x-ray is then injected, opacifying the coronary artery and allowing identification of coronary artery stenosis (narrowing) or blockage. Invasive coronary angiography is the current gold (or reference) standard for the detection of coronary stenosis or blockage. Most experts consider a narrowing of the coronary artery of more than 50% to be significant and a narrowing of more than 70% to warrant treatment with coronary revascularization (stenting or bypass). While major complications are rare, risks from coronary angiogram include heart attack, stroke, injury to the catheterized artery, irregular heart rhythms, allergic reaction to the contrast or medications used during the procedure, a tear in heart or artery, kidney damage, excessive bleeding, infection, blood clots and/or exposure to radiation from x-rays (Mayo Clinic 2011a).

Coronary computed tomographic angiography (CCTA): Computed tomographic coronary angiography is a noninvasive procedure used to determine whether obstructive CAD is present. Computed tomographic coronary angiography uses a CT scanner to produce cross-sectional x-ray images (slices) of the heart, which, in turn, are computer processed to create three dimensional images of the heart and its arteries. Similarly to ICA, CCTA uses an intravenous contrast agent that opacifies the coronary arteries permitting analysis of their lumens and detection of stenosis or blockage. Although it is noninvasive, risks of CCTA can include allergic and toxic reactions to contrast material and exposure to radiation (Mayo Clinic 2011b). As one of several intermediate tests to diagnose CAD, CCTA may be used in place of (or in addition to) stress ECG, stress NPI and stress echocardiography. *Coronary computed tomographic angiography produces anatomic information about the coronary arteries whereas the other non-invasive tests provide functional information about the myocardium, heart muscle, (presumably due to reductions in coronary artery flow or ischemia). In some diagnostic settings, functional information may be more valuable than anatomic information from CCTA and vice versa (*Institute for Clinical and Economic Review [ICER] 2008, Ollendorf 2010).

Computed tomographic coronary angiography has rapidly evolved over the past ten years. The coronary arteries which measure less than one centimeter in diameter are situated on the surface of the heart. The heart is moving both with respiration and with cardiac contractions. As such, the coronary arteries are moving in complex directions over the period of fractions of a second. The small size and complex motion of the coronary arteries require the acquisition of CT images of high spatial and temporal resolution. Multi-slice CT scanners obtain more than one slice per scanning cycle; current scanners obtain 64, 128 or 256 slices per cycle. This means that at least 64 slices can be obtained per cycle (e.g., every one second). The slices can

therefore be as narrow as one millimeter thick and the entire heart can still be scanned in ten to fifteen seconds. By gating image acquisition only during the diastolic portions of the cardiac cycle (when motion of the heart is minimal), multi-slice scanners of at least 64 slices allow sufficient spatial and temporal resolution to produce diagnostic images of the coronary arteries. Of note, there are several patient conditions that *preclude* the performance of CCTA *or limit the quality* of CCTA, as listed below. The percentage of patients unable to have CCTA because of these conditions is not known; the local percentage would be important for anyone planning to use CCTA as the first diagnostic test (following the history, physical examination, EKG, and cardiac enzymes) for patients with low to intermediate risk of CAD.

- Contrast allergy or renal insufficiency prevents the use of contrast agents.
- Patient obesity limits the quality of the images.
- Inability to hold one's breath for at least 20 seconds results in excessive respiratory motion.
- Cardiac arrhythmias and heart rates of greater than 60 beats per minute make cardiac gating difficult and thus increase cardiac motion.
- Coronary artery calcifications create artifacts that interfere with analysis of the coronary artery lumens.

Policy context

The technological evolution of CCTA has resulted in substantial growth in its use. Findings from a 2009 survey indicate that *nearly half of US-based cardiology practices now own or lease cardiac CT equipment* (Ollendorf 2010). In addition, CCTA is currently being advocated in the following clinical settings:

- Evaluation of stable chest pain in patients with low to intermediate probability of CAD;
- Evaluation of patients presenting to the emergency department (ED) with chest pain; and no evidence of myocardial infarction on ECG or cardiac enzyme blood tests;
- Evaluation of patients with high probability of CAD and atypical clinical presentation or non-diagnostic stress ECG;
- Evaluation of symptomatic patients after cardiac stent placement; and
- Screening of symptomatic patients who have intermediate to high risk of CAD.

The total *costs for CCTA* range from $300 to $800 per procedure, with Medicare reimbursement rates at $508 per CCTA (Halpern 2010; Ladapo 2009). In comparison, ICA is estimated to cost $1750 to almost $3000 per procedure, with Medicare reimbursement rates of $2948 per cardiac catheterization (Halpern 2010; Ladapo 2009). Halpern (2010) broke down the Medicare fee schedules for both CCTA and cardiac catheterization (Table 4).

Table 4. Costs of CCTA and ICA

Procedure	CPT Codes	Medicare Fee (US $)
Coronary CT Angiography (CCTA)		
Professional fee	0146T	$186
Technical fee	0146T	$322
Total	0146T	$508

Procedure	CPT Codes	Medicare Fee (US $)
Cardiac catheterization (ICA) Professional fee Technical fee	93510, 93543, 93556, 93545	 $354 $2594
Total		$2948

Key Questions

1. **What is the sensitivity and specificity of CT angiography in diagnosing obstructive coronary artery disease compared to catheter-based angiography with or without clinical follow-up?**

2. **Are there patients, situations or setting where the results of CT angiography would preclude the use of catheter-based angiography without changing clinical outcomes?**

3. **What are the rates of revascularization procedures, hospitalizations and utilization of other diagnostic tests following CT angiography compared to catheter-based angiography?**

4. **What is the evidence for harms related to CT angiography compared to catheter-based angiography?**

Methods

Search strategy

A full search of the MED clinical evidence core sources was done to identify systematic reviews (SRs), meta-analyses (MAs), technology assessments (TAs), and clinical practice guidelines (CPG) using the terms computed tomographic coronary angiography, CCTA, and diagnosis. Searches of core sources were limited to citations which were published after 2000. The core sources searched included: Hayes, Inc., ECRI, Cochrane Library (Wiley Interscience), UK National Institute for Health and Clinical Excellence (NICE), Blue Cross/Blue Shield Health Technology Assessment (HTA) program, Veterans Administration TA program, BMJ Clinical Evidence, the Canadian Agency for Drugs and Technologies in Health (CADTH), Washington State HTA, U.S. Services Preventive Task Force, and the Agency for Health Research and Quality (AHRQ).

A MEDLINE (Ovid) search was conducted to identify SRs and MAs as well as additional diagnostic test studies published after June 2009 since a good quality SR of 89 studies assessing the diagnostic test characteristics of CCTA (e.g., sensitivity, specificity) was published in 2010 (Schuetz 2010). The ending search date for the Schuetz (2010) SR was June 2009. However, we searched for controlled trials and observational studies that examined the impact of using CCTA on clinical outcomes (e.g., myocardial infarction, revascularization procedures) and further diagnostic testing (e.g., stress ECG, invasive coronary angiography), hospitalization following evaluation in the emergency room, cost, and harms from 2005 (the first published studies of scanners with 64 slices) through February 2010. Please see Appendix A for the full MEDLINE search strategy.

A search for relevant CPGs from 2005 through 2010 was also conducted, using the following sources: the National Guidelines Clearinghouse database, the Institute for Clinical Systems Improvement (ICSI), the Scottish Intercollegiate Guidelines Network (SIGN), the National Institute for Health and Clinical Excellence (NICE), the Veterans Administration/Department of Defense (VA/DOD) guidelines, American Heart Association, and the American College of Chest Physicians.

Inclusion criteria

Population: Adults with chest pain and suspected obstructive coronary artery disease

Intervention: Coronary computed tomographic angiography (CCTA)

Comparator: Catheter-based invasive coronary angiography (ICA)

Outcome: Diagnosis of obstructive coronary artery disease (greater than or equal to 50% stenosis), prognosis (including need for revascularization procedures, hospitalizations, coronary heart disease events), and harms (complications of the procedure, radiation dose, acute kidney injury, incidental findings).

Setting: Emergency department, inpatient and outpatient settings

Exclusion criteria

Studies were excluded if they published:

- in languages other than English;
- before January 2009 (for diagnostic test studies); or
- before January 2005 (for controlled trials and observational studies assessing outcomes including utilization, cost and harms).

Quality assessment

The methodological quality of the included studies was assessed using standard instruments developed and adapted by the MED Project that are modifications of the systems in use by NICE and SIGN (Guyatt 2008; NICE 2009; SIGN 2009). All studies and guidelines were assessed by two independent and experienced raters. In cases where there was not agreement about the quality of the study or guideline the disagreement was resolved by conference or the use of a third rater.

Each study was assigned a rating of good, fair, poor, based on its adherence to recommended methods and potential for biases. In brief, good quality SRs included a clearly focused question, a literature search that was sufficiently rigorous to identify all relevant studies, criteria used to select studies for inclusion (e.g., randomized controlled trials (RCTs)) and assess study quality, and assessments of heterogeneity to determine if a meta-analysis would be appropriate. Good quality RCTs clearly described the population, setting, intervention and comparison groups; randomly allocated patients to study groups; concealed allocation; had low dropout rates; and reported intention-to-treat analyses. Good quality SRs and RCTs also had low potential for bias

from conflicts of interest and funding source. Good quality diagnostic test studies prospectively studied consecutive patients who are representative of patients who will receive the test in practice, performed the test and the reference (gold) standard on all patients with the decision to perform the test and reference standard being independent of each other, and interpreted the results of the test and reference standard in a blinded fashion (without knowledge of the results of the other test). All good quality studies included *conflict of interest statements and descriptions of the source of funding and guarded* against undue influence of these factors. Fair quality SRs, RCTs and other diagnostic test studies had incomplete information about methods that might mask important limitations. Poor quality SRs, RCTs and diagnostic test studies had clear flaws that could introduce significant bias.

The overall strength or quality of the evidence was rated using a modified version of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system. After assessing the quality of individual studies a summary judgment of the overall quality of evidence was made for each key question and/or outcome (Guyatt 2008). The GRADE system defines the quality of a body of evidence for an outcome in the following manner:

- High: Further research is *very unlikely* to change our confidence in the estimate of effect. Typical sets of studies would be large RCTs without serious limitations.
- Moderate: Further research is *likely* to have an important impact on our confidence in the estimate of effect and may change the estimate. Typical sets of studies would be RCTs with some limitations or well-performed observational studies with additional strengths that guard against potential bias and have large estimates of effects.
- Low[1]: Further research is *very likely* to have an important impact on our confidence in the estimate of effect and is likely to change the estimate. Typical sets of studies would be RCTs with very serious limitations or observational studies without special strengths.

The methodological quality of the guidelines was assessed using an instrument adapted from the Appraisal of Guidelines Research and Evaluation (AGREE) Collaboration (AGREE 2009). Each guideline was assigned a rating of good, fair, poor, based on its adherence to recommended methods and potential for biases. A good guideline has fulfilled all or most of the criteria. A fair quality guideline will have fulfilled some of the criteria and those criteria not fulfilled are thought to unlikely alter the recommendations. If no or few of the criteria have been met, the guideline should be rated as poor. All guidelines were assessed by two independent and experienced raters. In cases where there was not agreement about the quality of the study or guideline the disagreement was resolved by conference or the use of a third rater.

Findings

Our MED Project core source search identified three SRs, 13 TAs, one RCT and five CPGs relevant to this topic. Because the Schuetz (2010) and Ollendorf (2010) SRs were good quality

[1] The MED Project collapses the low and very low GRADE categories because they usually have the same policy implications.

and included almost all of the diagnostic test studies included in prior SRs and TAs, we included only these two SRs. We also included the Washington State Health Care Authority Health Technology Assessment (HTA) (ICER 2008) because 1) it provided a good quality SR; 2) it focused on the clinical decision making related to the use of CCTA; and 3) it included studies of clinical outcomes, harms and costs. The MEDLINE search retrieved 747 full citations. After a review of citations and abstracts, we identified two additional diagnostic test studies (Hamirami 2010; Weustink 2010) and one RCT (Goldstein 2007) and four additional observational studies (Abidov 2009; Hadamitzky 2009; Karlsberg 2010; May 2009), not included in the SRs, that addressed clinical outcomes, harms or costs. A total of five guidelines were identified and used in this report. Detailed evidence tables by Key Question are presented in the appendices.

Key Question #1: What is the sensitivity and specificity of CT angiography in diagnosing obstructive coronary artery disease compared to catheter-based angiography with or without clinical follow-up?

Systematic reviews

Three recent good quality SRs with meta-analyses (ICER 2008; Ollendorf 2010; Schuetz 2010) examined the diagnostic accuracy of CCTA for the detection of CAD compared to ICA as a reference standard (Appendix B). Schuetz (2010) included 89 prospective studies using CT scanners of at least 12 detector rows (slices per scan cycle). *Their meta-analysis gave a pooled mean sensitivity of 97% (95% confidence interval [CI], 96%-98%), specificity of 87% (95% CI, 85%-90%), likelihood ratio (LR)*[2] *associated with a positive test result of 7.7 (95% CI, 6.2-9.5) and likelihood ratio associated with a negative test result of 0.03 (95% CI, 0.02-0.04).* They also found that scanners with at least 16 slices compared to scanners with fewer slices had a statistically significant increase in sensitivity (98% vs. 96%, respectively, $p < 0.05$) but not in specificity (90% vs. 85%, respectively, $p = 0.07$). A patient heart rate under 60 beats per minute was also associated with a statistically significant increase in sensitivity (99% vs. 96%, respectively, $p < 0.001$) but not specificity (86% vs. 87%, respectively, $p = 0.55$). Given the very high sensitivity (97%) and reasonably high specificity (87%), Schuetz (2010) concluded that CCTA can accurately detect and rule out CAD (stenosis greater than or equal to 50%) when compared to ICA, the current reference or gold standard.

It should be noted that the prevalence of CAD in the studies analyzed by Schuetz is quite high (59%) compared to the population who would likely get this test and may result in an overestimate of both sensitivity and specificity. Schuetz (2010) rated the articles for quality and determined them to be poor to fair in quality. Because so many of the patients in these studies had CAD, Schuetz (2010) noted these studies had a moderate to high risk of disease spectrum and verification bias. These biases would result in an overestimation of the sensitivity and specificity of CCTA.

[2] Likelihood ratio (LR) is the likelihood or odds that a given test result would occur in a patient *with the disease* compared to the likelihood that the same test result would occur in a patient *without the disease*. Likelihood ratios greater than 1.0 increase the probability of disease, and LRs less than 1.0 decrease the probability of disease. Likelihood ratios have a large impact on the probability of disease when they are greater than 10 or less than 0.1.

The SR by Ollendorf (2010) included a meta-analysis of 42 studies comparing the diagnostic accuracy of CCTA for 64-slice CT scanners to ICA. Ollendorf (2010) included fewer studies than Schuetz (2010) because Ollendorf (2010) restricted their SR to CT scanners with 64 slices or greater. Pooled sensitivity and specificity are similar to Schuetz (2010), sensitivity 98% (95% CI, 96%-99%) and specificity 85% (95% CI, 81%-89%).

The technology assessment performed by ICER (2008) for the Washington State Health Care Authority included a meta-analysis of 41 studies comparing 64-slice CCTA to ICA, the reference standard. Pooled sensitivity and specificity of CCTA for detection of CAD are similar to Schuetz (2010) and Ollendorf (2010), sensitivity 98% (95% CI, 97%-98%) and specificity 82% (95% CI, 80%-84%).

Diagnostic test studies (cross-sectional)

All studies identified since 2009 evaluated only 64-slice CT scanners. Two recent cross-sectional diagnostic test studies (Harnirani 2010; Weustink 2010) evaluated the diagnostic accuracy of 64-slice CCTA for the detection of CAD using ICA as a reference standard. Weustink (2010), a fair quality study, evaluated 517 consecutive patients referred for ICA in the Netherlands. Comparison of stress ECG and CCTA for diagnostic accuracy showed better sensitivity, specificity, positive likelihood ratio and negative likelihood ratio for CCTA compared to stress ECG in the overall group and in subgroups based on pre-test probability of CAD, defined as low (less than 20%) intermediate (20-80%) and high (greater than 80%). The overall sensitivity was 99% (95% CI, 97%-100%) and specificity 89% (95% CI, 84%-93%). Specificity varied from 72% (pre-test probability greater than 80%) to 93% (intermediate pre-test probability of 20%-80%) depending on the pre-test probability. In contrast, the sensitivity was stable across pre-test probabilities (99%-100%). The complete results are given in Appendix B.

Hamirami (2010) retrospectively reviewed 122 patients with intermediate to high pre-test probability of CAD who underwent CCTA <u>and</u> stress nuclear perfusion imaging (NPI) prior to ICA (Appendix B). For patients with coronary stenosis of more than 50%, CCTA had higher sensitivity and specificity than NPI (sensitivity, 99% vs. 56%, respectively; specificity, 74% vs. 39%, respectively).

Overall summary, quality and limitations of the evidence

There is consistency across studies of the diagnostic accuracy of CCTA compared to ICA for the detection of CAD. *The sensitivity is 97% or better and the specificity ranges from 72%-93%*. Coronary computed tomographic angiography may be better than stress ECG or NPI in diagnosing CAD.

The studies are all cross-sectional and employ convenience samples using patients already scheduled for ICA. This method of sampling patients creates a bias toward enrolling patients with a higher prevalence of CAD (mean prevalence 59%) than that expected in patients likely to get CCTA. Many patients who are likely to get a CCTA are in emergency room or primary care settings and would have a lower probability of CAD, versus a 59% probability or prevalence of CAD. Estimates of the sensitivity and specificity of CCTA from these studies may be higher than estimates from samples of patients who would get the test, creating a spectrum bias or effect.

Although many of the studies were poor to fair quality, the overall quality of the evidence is moderate because of the consistency of the results across the studies. However, the estimates of sensitivity and specificity should be viewed as the upper bound of sensitivity and specificity due to spectrum bias (from higher prevalence of CAD in the studies compared to the prevalence in patients who might typically get CCTA).

Key Question #2: Are there patients, situations or settings where the results of CT angiography would preclude the use of catheter-based angiography without changing clinical outcomes?

Systematic reviews

The Washington State Health Technology Assessment (ICER 2008) evaluated studies that assessed the impact of CCTA on patient outcomes (Appendix C). ICER (2008) identified seven studies that address patient outcomes, all coming from the ED setting. The evidence for this Key Question is based on one fair quality RCT (Goldstein 2007) and six poor quality case series (Hoffmann 2006; Hollander 2007; Johnson 2007; Pundziute 2007; Rubenshtein 2007; Savino 2006). Goldstein (2007) randomly assigned 197 patients without known CAD seen in the ED for chest pain to CCTA or usual care. All patients had normal ECGs and cardiac enzymes, so were at very low to low risk for myocardial infarction. Of the 99 patients randomized to the CCTA care arm, 67 (68%) patients with normal CCTA were discharged home rapidly per protocol. Twenty-four (24%) patients with indeterminate or non-diagnostic CCTA had stress NPI, and 8 (8%) patients with severe CAD on CCTA had ICA. The 98 patients in the control arm had standard ED observation, serial cardiac enzymes and stress NPI. Of these, 93 (95%) had normal stress NPI and were discharged home. No patients in either arm of the study had adverse outcomes (i.e., death, myocardial infarction, unstable angina, or test complications) at six months follow-up. In this small study, there were savings of evaluation time and money due to earlier discharge of patients without evidence of CAD. Six poor quality case series also demonstrated that a negative CCTA allows discharge patients from the ED without adverse events (Hoffmann 2006; Hollander 2007; Johnson 2007; Pundziute 2007; Rubenshtein 2007; Savino 2006). ICER (2008) applied these data to 1000 hypothetical 55 year old men seen in the ED with chest pain and normal ECGs, and initial cardiac enzymes. *Out of 1000 patients, early CCTA would result in 456 patients being discharged early, a reduction of false negative diagnoses for CAD from 51 to 5, a reduction in ICAs from 464 to 380 and a reduction of negative ICAs from 246 to 116 (ICER 2008).*

The Washington HTA cautions that the studies are all small and that the rate of acute coronary syndrome (ACS) and cardiac events were quite low; this low prevalence could make the strategy of early discharge seem better than it would in a patient setting with higher risk of CAD and ACS.

The Ollendorf (2010) SR included one RCT (Goldstein 2007) and six case series from ED settings and four case series from outpatient settings (Appendix C); all but three of the case series (Danciu 2007; Hay 2009; Wagdi 2009) were included in the ICER (2008) report. The seven ED studies found that early triage of low risk patients with CCTA resulted in no adverse cardiac events and was time saving compared to standard ED care. Four case series of CCTA for low risk patients with chest pain in the outpatient setting found no adverse cardiac events over 12

to 19 months of clinical follow-up. Ollendorf (2010) found serious methodological flaws in almost all of the studies evaluated.

RCTs

The only fair quality RCT (Goldstein 2007) was included in the two reviews above.

Other study designs

We found one good quality prospective cohort study and two poor quality cohort or case series studies not included in the Washington HTA (ICER 2008) or Ollendorf SR (2010) (Appendix C).

Hadamitzky (2009), a good quality cohort study, prospectively followed 1150 patients referred for CCTA in Germany. Of 802 (70%) patients without CAD on CCTA, there were four major adverse cardiac events (MACEs) (0.5%) and one non-cardiac death (0.1%). For 348 (30%) patients with CAD on CCTA, there were 17 cardiac events (5%) and six non-cardiac deaths (2%). In this study, findings on CCTA predicted the subsequent occurrences of MACE and non-cardiac deaths with statistical significance ($p < 0.001$ and $p = 0.004$, respectively).

May (2009), a poor quality case series, compared length of stay and total expense for 53 consecutive ED patients with chest pain and low probability of CAD. Three patient groups were compared: standard care with observation and serial enzyme determinations, CCTA with further observation and CCTA with discharge if the CCTA was negative for CAD. Length of stay in the ED was reduced from 25 hours for standard care to 14 hours for CCTA and observation to 5 hours for CCTA with discharge ($p < 0.001$ between standard care and CCTA with discharge). Total costs were reduced from $7600 for standard care to $6100 for CCTA with observation to $4250 for CCTA with discharge ($p < 0.001$ between standard care and CCTA with discharge). No adverse cardiac events were reported in any of the groups on follow-up at six months.

Abidov (2009), a poor quality case series, prospectively followed 199 patients referred for CCTA after a non-diagnostic NPI study. For the 93 patients that did not have CAD on CCTA, there were no cases of subsequent ICA or MACE after two years of follow-up. Of the 36 patients with 70% coronary stenosis, there were 24 (67%) subsequent ICAs and 12 MACEs.

Overall summary, quality, strengths and limitations of the evidence

The ICER (2008) and Ollendorf (2010) SRs included the only good quality RCT. In addition a number of retrospective and prospective cohort studies and case series have examined the rate of MACE, ICA or percutaneous coronary interventions (PCI) on follow-up of six months to three years. *The results are consistent both in the ED and in the outpatient setting that with a low to intermediate pre-test probability of CAD, a negative CCTA result is reliable in predicting the absence of subsequent cardiac events.* Although many of the observational studies are poor to fair quality, the overall quality of the evidence is moderate based on data from one fair quality RCT, one large good quality cohort study (Hadamitzky 2009), and the consistency of results across all studies.

Key Question #3: What are the rates of revascularization procedures, hospitalizations and utilization of other diagnostic tests following CT angiography compared to catheter-based angiography?

We found no studies that assessed revascularization or hospitalization rates following CCTA compared to ICA.

Other study designs

One poor quality, before and after study addressed the utilization of additional diagnostic testing following introduction of CCTA into a single cardiology practice (Appendix D). Karlsberg (2010) looked at the utilization of stress ECG, stress NPI and ICA in a large university and community based cardiology practice in southern California for the two years prior and two years following the introduction of CCTA into their practice. Review of the overall demographics of the patient base and other measures of the practice suggests that there was no significant change in the characteristics of the practice or its patients during the four years of observation. After the introduction of CCTA, the number of stress ECGs did not change but the number of stress NPIs decreased by 19% compared to before CCTA ($p = 0.02$). The number of ICAs decreased by 47% ($p < 0.01$) and the yield of positive results on ICA (i.e., ICA results led to PCI) increased from 19% to 28% ($p < 0.001$). The results suggest that CCTA may decrease the use of other intermediate diagnostic tests (i.e., stress NPI) and improve the selection of patients for ICA.

Overall summary, strengths and limitations of the evidence

Low quality, limited evidence suggests that findings on CCTA may decrease the use of other diagnostic tests and ICA.

Key Question #4: What is the evidence for harms related to CT angiography compared to catheter-based angiography?

Systematic review

The Washington HTA (ICER 2008) addresses the issue of harms associated with CCTA including radiation dosage, risks of contrast agents and the detection of unrelated findings on CT scans (incidental findings). Washington HTA concluded that CCTA is safe. Reported rates of serious contrast reactions and contrast induced nephropathy have been very low; the type and amount of contrast used is identical to other common CT procedures including ICA.

The amount of radiation dose for CCTA is similar to a CT scan of the abdomen or an ICA. It is not possible to measure absorbed dose for any patient directly and the amount of radiation absorbed varies with patient size and body type. Radiation dose is usually estimated based on radiation phantom studies. Table 5, which is adapted from ICER (2008), shows relative amounts of radiation from various exposures.

Table 5: Relative Amounts of Radiation Exposure from Different Sources

Radiation exposure scenario	Approximate effective dose (mSv)	Approximate dose in chest x-ray equivalents
Chest x-ray	0.02	1

Radiation exposure scenario	Approximate effective dose (mSv)	Approximate dose in chest x-ray equivalents
Round trip airline flight Seattle to New York	0.06	3
Computed Tomography Head	2.0	100
Computed Tomography Abdomen	10.0	500
CCTA	8-14	400-700
CCTA with newer gating techniques	2-5	100-250
Invasive coronary angiography (ICA)	5-7	250-350
Nuclear perfusion imaging (NPI)	9-13	450-650

Of note, CCTA followed by NPI followed by ICA may result in up to 1700 chest x-ray equivalents of radiation dose. Using CCTA to replace NPI is a more reasonable strategy than adding it to NPI. Hopefully, replacing NPI with CCTA will decreased the use of ICA and result in less radiation exposure overall.

There are relative benefits and harms from the incidental findings noted on CT of the chest (findings in the chest obtained during a CCTA). Approximately 40 to 80% of patients undergoing CCTA will have a finding that is not related to the coronary arteries; five to twenty percent will have a finding deemed clinically important enough for further evaluation. Although some of the patients with these incidental findings will have been judged to have received some benefit, findings from the few studies that have examined this question suggest that the proportion of patients receiving some benefit is very low, while additional risks, anxieties and costs are generated by the additional investigations.

RCTs and other study designs

There are no other studies that address this Key Question.

Overall summary, strengths and limitations of the evidence

The evidence for this Key Question comes from the Washington HTA (ICER 2008). The evidence on contrast reactions, radiation dose and incidental findings is well established from epidemiologic and other studies.

Guidelines

Summary of Guidelines and Quality Assessment

Five guidelines address one or more of the Key Questions (Appendix E). One guideline from the *Scottish Intercollegiate Guidelines Network (SIGN)* (2007) addresses CCTA only peripherally. This guideline concludes that CCTA is effective in establishing a diagnosis of CAD when performed by trained and skilled teams.

The *American College of Radiology (ACR) Appropriateness Criteria* (ACR 2008, 2010a, 2010b) address the appropriateness of CCTA in several different clinical scenarios. These include chest pain suggestive of ACS, acute chest pain with low probability of CAD and chronic chest pain with high probability of CAD. The ACR guidelines point out that CCTA may be limited in patients with 1) heart rates more than 65 beats per minute; 2) arrhythmias; and 3) dense calcifications of the coronary arteries. The guidelines also comment that in patients with high probability of CAD, a negative CCTA may be a false negative. In these patients ICA may be more appropriate than CCTA.

A joint guideline from the *American College of Cardiology Foundation* and a number of other organizations (Taylor 2010) also gives appropriateness ratings for CCTA. Clinical scenarios for which CCTA is thought appropriate include 1) acute chest pain with suspicion of ACS and low to intermediate probability of CAD (ED setting); 2) non-acute symptoms possible representing ischemic equivalent but with normal, uninterpretable or non-diagnostic ECG; 3) risk assessment (screening) in asymptomatic patients without known CAD with known family history and low pre-test probability or with intermediate pre-test probability and no known family history; and 4) in patients with continued symptoms and a prior normal stress ECG. Of note, other guidelines do not recommend CCTA for asymptomatic patients (i.e., as a screening test).

The *American College of Cardiology Foundation* (Kramer 2007) also issued a statement of competence for physicians wishing to perform CCTA. This guidelines recommends that physicians who perform CCTA should have knowledge of CT hardware, competence in acquisition and interpretation techniques, knowledge of post-processing (including the use of computer work stations), knowledge in contrast reactions and their treatment, knowledge of radiation, and techniques to reduce radiation exposure. The guidelines recommend fellowship training and a minimum of 50 cases per year and 30 hours of CME training every three years.

Comparison of guidelines and evidence summary

The guidelines are poor to fair quality. All of the guidelines recommend the use of CCTA for evaluation of patients without known CAD who present with acute chest pain and have low to intermediate risk of CAD. This recommendation is consistent with the evidence in Key Questions 1 and 2.

Recommendations for use of CCTA in asymptomatic patients would result in a substantial proportion of patients having a false positive test result, based on the evidence of diagnostic accuracy of CCTA in Key Question 1. At a 15% pre-test probability or prevalence of obstructive CAD in an asymptomatic population, the post-test probability of CAD with a positive CCTA (assuming a sensitivity of 97% and specificity of 85%) would be 54% (Appendix F). Almost half (46%) of patients with a positive CCTA would have a false positive test resulting in further evaluation and/or treatment. Even with a 25% pre-test probability (prevalence) of CAD, a positive result on a CCTA would result in a post-test probability of 68%, and 32% of patients would have a false positive test result. Although we did not find studies that addressed screening asymptomatic patients, we did not specifically search for evidence for the use of CCTA as a screening test because it was beyond the scope of this report. The HTA by ICER

(2008) identified only one study of CCTA as a screening test in asymptomatic patients with diabetes. The anticipated completion date for this study is December 2011.

Summary

General conclusions

Coronary computed tomographic angiography is an intermediate diagnostic test used to increase (if the test is positive) or decrease (if the test is negative) the probability of CAD in patients presenting with chest pain. It is used as an intermediate diagnostic test similar to other intermediate tests such as stress ECG, stress NPI and stress ECHO. In contrast to the other intermediate tests, *CCTA provides anatomic information about the coronary arteries, but not functional information about the myocardium* (heart muscle function).

Patient and technical factors affect the use and quality of CCTA. These factors are reflected in study inclusion and exclusion criteria and in guideline recommendations. The factors relate to 1) choice of patients; 2) choice of CT scanner; and 3) competence of physicians performing the procedure. Patient conditions that *preclude* the performance of CCTA *or limit the quality* of CCTA include:

- Obesity;
- Inability to hold one's breath for at least 20 seconds;
- Cardiac arrhythmias and heart rates of greater than 60 beats per minute;
- Coronary artery calcifications creating artifacts that interfere with analysis; and
- Contrast allergy or renal insufficiency.

All recent studies evaluate 64-slice CT scanners, and most experts state that 64 slices is the current minimum number of slices needed to perform the study with acceptable spatial and temporal resolution. Finally, physicians performing CCTA should have additional training in all aspects of the procedure and a minimum of 50 cases per year is recommended to maintain competence in the procedure.

Coronary computed tomographic angiography has a very high sensitivity (98%) and moderate to moderately high specificity (72-93%) for the detection of coronary artery stenosis, based on moderate quality evidence. These performance characteristics support the use of CCTA to "rule out" obstructive CAD in emergency room patients with acute chest pain and normal ECGs and initial cardiac enzymes and in outpatients with stable chest pain. The major utility of CCTA in these settings is the ability to identify those patients with no CAD, so they can be safely discharged without additional work-up or concern for future cardiac events.

In patients with low to intermediate risk of CAD, CCTA appears to have better diagnostic accuracy than ECGs and NPI, based on low to moderate quality evidence. In the ED setting, use of CCTA may reduce the length of stay and total costs. This is substantiated by one small randomized controlled trial (n = 197) and 7 observational studies suggesting that emergency room patients with low to intermediate pre-test probability of CAD and who have a negative CCTA may be discharged quickly without concern for MACE. Finally, CCTA may reduce the

number of subsequent tests including stress NPI and ICA, based on a single poor quality before and after study.

Although two specialty-based guidelines find CCTA "appropriate" for patients with chest pain and high risk of CAD and for screening of high risk asymptomatic patients, other guidelines recommend against these uses. Few studies specifically address these uses and the pre-test probabilities of CAD in these patient groups would likely result in a high number of false negatives (in symptomatic patients at high risk of CAD) and false positives (in asymptomatic patients at low to intermediate risk of CAD). For example, screening 10,000 asymptomatic patients with a 20% prevalence of obstructive CAD with CCTA would result in 1960 patients with a true positive test result and almost as many (1200 patients) with a false positive test result (Appendix F). This would result in 1200 patients potentially receiving unnecessary testing (e.g., ICA) and treatment.

Limitations of the evidence

The evidence on diagnostic accuracy comes primarily from convenience samples of patients already scheduled for ICA, increasing the prevalence or pre-test probability of CAD in these samples. This may lead to an over-estimation of the sensitivity and specificity of CCTA than that which would be expected based in a more typical population of patients likely to get CCTA. The evidence on the effect of CCTA on subsequent cardiac diagnostic testing and on subsequent cardiac events is based on only one small fair quality RCT, one good quality cohort study, 12 poor to fair quality cohort or case series studies, and one before and after study.

Policy Considerations

This report reviews current Medicare, private payor and participating MED state policies regarding CCTA coverage. A total of 13 policies were reviewed with respect to six key elements identified in the evidence review:

1. Use of low to intermediate patient pre-test probability of CAD to determine coverage;
2. Other patient criteria (e.g., obesity, heart rate, ability to hold one's breath);
3. Exclusion of coverage for patients with known coronary artery disease;
4. Exclusion of coverage for screening in asymptomatic patients;
5. Requirement to use a 64-slice scanner; and
6. Requirement regarding physician competence to perform CCTA.

The 13 policies include nine Medicare Local Coverage Determinations (LCDs) applicable to participating MED states, two private payor policies (Aetna, Cigna), and two state policies (Minnesota, Washington). In 2008, Medicare determined that it would not issue a national coverage determination with respect to CCTA and that coverage should be determined by local contractors (CMS 2008). We therefore focus our review on LCDs applicable to MED participating states, which comprise a total of nine current LCDs. In addition, this analysis focuses on policy criteria relevant to the use of CCTA for purposes of diagnosing CAD, and does not address CCTA coverage for other purposes.

Appendix G provides a table summarizing the analysis of the 13 policies according to the six elements above. Appendix H provides further detail regarding coverage criteria for each policy.

1. Patient Pre-test Probability of CAD Criteria

With respect to use of CCTA for diagnostic purposes, four of the 13 policies reviewed limit coverage of CCTA exclusively to patients with low or intermediate pre-test probabilities of CAD (Aetna, Cigna, Medicare LCD for WV, Washington). These policies support the use of CCTA as a test to "rule out" obstructive CAD in patients with low to intermediate pre-test probabilities of CAD, and are consistent with the evidence findings in this report. Use of CCTA for these purposes is most likely to occur in ED and primary care settings.

We also note that several of the Medicare LCDs contain confusing language with respect to the use of high patient pre-test probability criteria in order to discourage use of CCTA for screening purposes (Medicare LCD for AR and OK). We highlight this point as a general caution for states using Medicare LCDs as a basis or reference for developing state CCTA coverage criteria.

2. Other Patient Criteria

Five of the 13 policies reviewed identify other patient criteria for coverage of CCTA, such as allergies, heart rate and obesity. Aetna's policy provides good example language and is consistent with the evidence findings of this report. Specifically, the policy states that CCTA is considered experimental and investigational for persons with any of the following contraindications to the procedure:

1. Body mass index (BMI) greater than 40;
2. Inability to image at desired heart rate (under 80 beats per minute), despite beta blocker administration;
3. Person with allergy or intolerance to iodinated contrast material;
4. Persons in atrial fibrillation or with other significant arrhythmia; or
5. Persons with extensive coronary calcification by plain film or with prior Angston score greater than 1700.

Four of the nine LCDs (AR (Part A and B), OK, WV) also contain patient criteria, although the language in several of these policies is unclear (AR, OK) and less comprehensive than the Aetna policy (WV).

3. Patients with Known Coronary Artery Disease

While none of the state or private payor policies cover CCTA for patients with known CAD, five of the nine LCDs reviewed cover CCTA for patients with known CAD, usually in patients who are recently post-stent or post-coronary artery bypass surgery and having recurrent symptoms. These five LCDs apply to coverage policies in four states (AL, AR, OK, WV). While this report does not review evidence in relation to use of CCTA in patients with known CAD, we include this coverage element in our analysis to acknowledge that several policies do not cover CCTA for patients with known CAD.

4. *Screening in Asymptomatic Patients*

All of the policies reviewed exclude coverage of CCTA for screening in asymptomatic patients.

5. *Requirement of 64-Slice Scanner*

Nine of the 13 policies reviewed require use of a minimum of 64-slice scanners. These include five of the nine LCDs (AK, MN, OR, WA, WV), both state policies (MN, WA), and both private payor policies (Aetna, Cigna) reviewed. One additional LCD recommends but does not require the use of 64-slice scanners (AL). The three policies that do not require 64-slice scanners are LCDs that apply to two states (AR, OK).

6. *Physician Competence to Perform CCTA*

All of the LCDs set forth expectations that CCTA tests are performed under the direct supervision of physicians who meet competency guidelines defined by the American College of Cardiology (ACC) and American Heart Association (AHA) Clinical Competence Statement on Cardiac Imaging with Computed Tomography and Magnetic Resonance (2005) and the American College of Radiology (ACR) Clinical Statement on Noninvasive Cardiac Imaging (2005). The state policies (MN, WV) and private payor policies (Aetna, Cigna) reviewed do not outline expectations of physician competence. Please see Appendix G for a summary analysis of each policy according to the six elements discussed above, and Appendix H for relevant coverage criteria excerpted from each policy.

Appendix A. Updated Search Strategy

1 exp Coronary Angiography/
2 exp Angiography/
3 exp Heart Diseases/di, ra [Diagnosis, Radiography]
4 exp Heart/
5 3 or 4
6 2 and 5
7 1 or 6
8 exp Tomography, X-Ray Computed/
9 7 and 8
10 exp "Sensitivity and Specificity"/
11 9 and 10
12 exp Catheterization/
13 11 and 12
14 Comparative Study/
15 9 and 12 and 14
16 13 or 15
17 (catheter$ adj7 ((cat or ct or compute$ or tomogra$) adj3 scan$)).mp. [mp=protocol supplementary concept, rare disease supplementary concept, title, original title, abstract, name of substance word, subject heading word, unique identifier]
18 5 and 17
19 16 or 18
20 exp coronary Angiography/ae, co, ct, mo [Adverse Effects, Complications, Contraindications, Mortality]
21 exp Tomography, X-Ray Computed/ae, ct, mo [Adverse Effects, Contraindications, Mortality]
22 exp Catheterization/co, ct, mo [Complications, Contraindications, Mortality]
23 20 or 21 or 22
24 5 and 23
25 limit 24 to (english language and yr="2009 -Current")
26 exp Radiation Dosage/
27 1 or 3
28 26 and 27
29 ae.fs.
30 ct.fs.
31 29 or 30
32 28 and 31

33 exp treatment outcome/
34 exp Myocardial Revascularization/
35 9 and 33
36 9 and 34
37 exp Hospitalization/
38 9 and 37
39 35 or 36 or 38
40 limit 39 to yr="2009 -Current"
41 limit 40 to english language
42 limit 19 to (english language and yr="2009 -Current")
43 limit 32 to yr="2009 -Current"
44 25 or 41 or 42 or 43

Appendix B. Sensitivity, Specificity, and Likelihood Ratios[3] for CT Angiography in Diagnosing Obstructive Coronary Artery Disease Compared to Catheter-Based Angiography with or without Clinical Follow-up

Reference Study Design and Number of Studies	Intervention(s); Comparator(s)	Outcomes Evaluated and Main Findings	Quality Rating* and Comments
Systematic Reviews and HTAs			
Schuetz (2010) SR and MA of 89 prospective diagnostic test studies	CCTA with CT scanners with at least 12 detector rows; >50% stenosis used as an abnormal finding for CAD. ICA used as a reference standard.	**Pooled Diagnostic Accuracy Mean (95% CI):** **Sensitivity / Specificity / Positive Likelihood ratio (LR) / Negative LR** Overall: 97% (96-98) / 87% (84-90) / 7.7 (6.2-9.5) / 0.03 (0.02-0.04) Suspect CAD: 98% (96-99) / 89% (86-92) / 9.1 (7.0-11.8) / 0.03 (0.02-0.04) Suspect ACS: 96% (87-98) / 77% 951-91) / 4.1 (2.0-8.4) / 0.06 (0.02-0.19) Heart rate < 60: 99% (98-99) / 86% (79-91) Heart rate > 60: 96% (95-97) / 88 (84-91)	Good quality SR and MA. Quality of the 89 studies was rated as poor to moderate by SR authors
Ollendorf (2010) SR and MA of 42 prospective diagnostic test studies	CCTA with CT scanners of 64 detector rows; 50% stenosis used as abnormal finding for CAD. ICA used as reference standard.	**Pooled Diagnostic Accuracy Mean (95 % CI)** **Sensitivity / Specificity** Overall: 98% (96-99%) / 85% (81-89%)	Good quality SR and MA. Publication bias, spectrum bias, clinical review bias noted in included studies.
ICER (2008) TA of 41 studies	CCTA compared to Stress NPI and Stress Echo; ICA as reference standard	**Diagnostic accuracy of CCTA in outpatient setting** Pooled sensitivity = 0.98 (95% CI 0.97 to 0.98) Pooled specificity = 0.82 (95% CI 0.80 to 0.84)	Good quality SR
Individual Studies			
Weustink (2010) Prospective case series; 517 consecutive patients	CCTA and stress ECG performed in all patients	**Diagnostic accuracy of stress testing and CCTA compared with ICA as a reference standard:** **Test / Sens / Spec / PPV / NPV / LR+ / LR-** *Overall* Stress ECG: 78% / 77% / 80% / 76% / 3.4 / 0.28 CCTA: 99% / 89% / 91% / 99% / 9.2 / 0.01	Fair quality

[3] Likelihood ratio = is the likelihood or odds that a given test result would be expected in a patient *with the disease or condition* compared to the likelihood that the same test result would be expected in a patient *without the disease*. Likelihood ratios that are > 1.0 increase the probability of disease and LRs less than 1.0 decrease the probability of disease. Likelihood ratios have a large and more significant impact on the probability of disease when they are > 10 or < 0.1.

Reference Study Design and Number of Studies	Intervention(s); Comparator(s)	Outcomes Evaluated and Main Findings	Quality Rating* and Comments
		Test **Sens** **Spec** **PPV** **NPV** **LR+**[4] **LR** *Pre-test probability<20%* Stress ECG 71% 83% 50% 92% 4.1 0.35 CCTA 100% 89% 69% 100% 8.9 0.00 *Pre-test probability20-80%* Stress ECG 79% 77% 79% 77% 3.5 0.28 CCTA 99% 93% 94% 99% 14.6 0.01 *Pre-test probability >80%* Stress ECG 79% 60% 91% 37% 2.0 0.36 CCTA 99% 72% 95% 95% 3.6 0.01	
Hamirami (2010) Retrospective case series; 122 patients with intermediate to high probability of CAD	CCTA and NPI compared to ICA as reference standard.	**Detection of coronary artery stenosis > 50% on ICA:** **Sens** **Spec** **PPV** **NPV** CCTA 99% 74% 92% 96% NPI 56% 39% 73% 23% **Diagnosis of coronary artery stenosis > 70% on ICA**: **Sens** **Spec** **PPV** **NPV** CCTA 90% 86% 92% 83% NPI 58% 43% 64% 36%	Poor quality

SR indicates systematic review; MA meta-analysis; HTA health technology assessment; TA technology assessment, CT computed tomography; CCTA coronary computed tomography angiography; CI confidence interval; CAD coronary artery disease; ACS acute coronary syndrome (myocardial infarction or unstable angina; ICA invasive coronary angiography also catheter coronary angiography; ECG electrocardiogram; NPI nuclear perfusion imaging; SENS sensitivity; SPEC specificity; PPV positive predictive values; NPV negative predictive values; LR+ likelihood ratio associated with a positive test result; LR- likelihood ratio associated with a negative test result.

* Quality ratings for the systematic reviews and technology assessment are based on the MED rating instruments derived from SIGN and NICE: Good, fair, poor. The quality of the individual studies included in the technology assessment and systematic reviews are based on various systems used by the authors. Many of these systems were derived from the Cochrane Collaboration's rating methods for diagnostic test studies.

[4] Likelihood ratio = is the likelihood or odds that a given test result would be expected in a patient *with the disease or condition* compared to the likelihood that the same test result would be expected in a patient *without the disease*. Likelihood ratios that are > 1.0 increase the probability of disease and LRs less than 1.0 decrease the probability of disease. Likelihood ratios have a large and more significant impact on the probability of disease when they are > 10 or < 0.1.

Appendix C: Subgroups of Patients, Situations or Settings Where the Results of CCTA Would Preclude the Use of Catheter-based Angiography without Changing Clinical Outcomes.

Reference Study Design and number of studies	Intervention(s); Comparator(s)	Outcomes Evaluated and Main Findings	Quality Rating* and Comments
Systematic Reviews and HTAs			
ICER (2008) SR of 1 RCT and 3 case series in ED settings	CCTA vs. standard care	**CCTA Impact on Patient Management and/or Clinical Outcomes:** **Author / Year / Setting / Sample Size / CAD Risk / Follow-up / Major Findings** RCT Goldstein 2007 ED 197 very low 6 months No MACEs over 6 mo of follow-up; CCTA alone correctly diagnosed CAD in 75% Case series Rubinshtein 2007 ED 58 intermediate 12 months Led to discharge in 45%; no MACE Pundziute 2007 ED 100 intermediate 16 months MACE 0% in CCTA neg. MACE 30% in pos Hollander 2007 ED 54 low 30 days No MACE recorded; CCTA dx 4/6 of ACS Johnson 2007 ED 55 n/a ≥5 months CCTA correctly diagnosed CAD in 93% Savino 2007 ED 23 n/a 12 months ED visit only; all CCTA stenoses confirmed Hoffman 2006 ED 103 low 5.2 months Sens for ACS 100%; spec for ACS 82%	Good quality SR of poor to fair quality studies
Ollendorf (2010) SR of 11 patient outcome studies in ER and outpatient settings	CCTA vs. standard care	**CCTA Impact on Patient Management and/or Clinical Outcomes:** **Author / Year / Setting / Sample Size / CAD Risk / Follow-up / Major Findings** RCT Goldstein 2007 ED 197 very low 6 months No MACEs over 6 mo of follow-up; CCTA alone correctly diagnosed CAD in 75% Case series Gallagher 2007 ED 85 low 30 days No MACE recorded; CCTA dx 6/7 ACS Hoffman 2006 ED 103 low 5.2 months Sens for ACS 100%; spec for ACS 82% Hollander 2007 ED 54 low 30 days No MACE recorded; CCTA dx 4/6 of ACS Johnson 2007 ED 55 n/a ≥5 months CCTA correctly diagnosed CAD in 93% Rubinshtein 2007 ED 58 intermediate 12 months Led to discharge in 45%; no MACE Savino 2007 ED 23 n/a 12 months ED visit only; all CCTA stenoses confirmed Danciu 2007 OP 421 intermediate 12 months MACE 0.3% in low risk group; 70.5% in ICA Pundziute 2007 OP 100 intermediate 16 months MACE 0% in CCTA neg. MACE 30% in pos Hay 2009 OP 138 intermediate 20 months Mace 0% in CCTA neg. MACE 31% in pos Wagdi 2009 OP 347 low-intermed 16 months 90% reduction of "normal ICA	Good quality SR of poor to fair quality studies

Reference Study Design and number of studies	Intervention(s); Comparator(s)	Outcomes Evaluated and Main Findings					Quality Rating* and Comments
Individual Studies							
Hadamitzky (2009) Prospective cohort of 1256 consecutive outpatients; 1155 in final study group; Germany	CCTA Median clinical follow-up of 18 months	**Rates of MACE:**					Good quality
		CCTA Findings	**No CAD**	**CAD**	**p value**	**Odds Ratio**[5]	
		# of patients	802	348			
		Cardiac death	0	0			
		Non fatal MI	0	1	0.13		
		Unstable angina	1	4	0.03		
		Late revascularization	3	15	<0.001		
		All card events	4	17	<0.001	16	
		Non-cardiac death	1	6	0.004		
May (2009) Retrospective case series of 53 consecutive ED patients	CCTA added to standard care vs. standard care	**Costs and lengths of hospital stay:**					Poor quality
				Length of stay	**Total charges**		
		Standard care		25.4 ± 6.3 h	$7597 ± $2216		
		CCTA with observation		14.3 ± 5.0 h	$6153 ± $1196		
		CCTA with discharge if CCTA negative		5.0	$4251 ± 420		
Abidov (2009) Prospective case series of 199 patients referred for CCTA after non-diagnostic NPI	CCTA Clinical follow-up of 2 years	**CCTA Findings**	**# of Patients**	**# of patients with ICA**	**# of patients with MACE/PCI**		Fair quality
		No CAD	93	0	0		
		CAD > 50%	70	8	1		
		CAD > 70%	36	24	12		

SR indicates systematic review; MA meta-analysis; HTA health technology assessment; RCT randomized controlled trial; ED emergency department; CCTA coronary computed tomographic angiography; CAD coronary artery disease; MACE major adverse cardiac events; ACS acute coronary syndrome (myocardial infarction or unstable angina); ICA invasive coronary angiography also catheter coronary angiography; MI myocardial infarction; NPI nuclear perfusion imaging; PCI percutaneous coronary intervention.

*Quality ratings for the systematic reviews and technology assessment are based on the MED rating instruments derived from SIGN and NICE: Good, fair, poor. The quality of the individual studies included in the technology assessment and systematic reviews are based on various systems used by the authors. Many of these systems were derived from the Cochrane Collaboration's rating methods for diagnostic test studies.

[5] Odds Ratio = Odds Ratio = the ratio of two odds. An odds is the ratio of the probability of an event occurring versus the probability of the event not occurring (probability of event/[1-probability of event]) for a group. In this case, the odds of MACE in patients with CAD on CCTA are 17 times greater than the odds for patients with no CAD on CCTA.

Appendix D: Rates of Revascularization Procedures, Hospitalizations and Utilization of Other Diagnostic Tests Following CT Angiography Compared to Catheter-based Angiography.

Reference Study Design and number of studies	Intervention(s); Comparator(s)	Outcomes Evaluated and Main Findings						Quality Rating* and Comments
Individual Studies								
Karlsberg (2010); chart review from one cardiology practice in California	CCTA	**Rates of utilization of stress ECG, stress NPI and ICA before and after introduction of CCTA:**						Poor quality
			2004	**2005**	**2006**	**2007**	**p value** (2004 vs. 2007)	
		Practice demographics						
		Office visits	31,855	29,617	20,049	30,066	ns	
		Hospital admissions/consults	3,416	2,878	2,773	3,468	ns	
		Non Invasive Procedures						
		ECG	15,679	13,358	13,309	14,670	ns	
		SPECT	3,223	3,139	2,810	2,614	0.021	
		CCTA	0	74	1,405	945	n/a	
		Invasive Procedures						
		ICA	2,083	1,848	1,589	1,150	0.012	
		PCI	405	457	425	326	ns	
		PCI/ICA	19%	25%	27%	28%	0.008	

CCTA indicates coronary computed tomography angiography; ECG electrocardiogram; stress refers to tests that are performed after a period of exercise on a treadmill; NPI nuclear profusion imaging; ICA invasive coronary angiography also catheter coronary angiography; SPECT single proton emission computed tomography; PCI percutaneous coronary intervention; ns not statistically significant; n/a not available.

*Quality ratings for the systematic reviews and technology assessment are based on the MED rating instruments derived from SIGN and NICE: Good, fair, poor. The quality of the individual studies included in the technology assessment and systematic reviews are based on various systems used by the authors. Many of these systems were derived from the Cochrane Collaboration's rating methods for diagnostic test studies.

Appendix E. Summary of Guidelines

Recommending body, year published	Guideline(s)	Evidence base	Overall quality*
Scottish Intercollegiate Guideline Network, Management of Stable Angina (SIGN 2007)	This older guideline only addresses CCTA peripherally. It states that CCTA is effective in establishing a diagnosis of coronary heart disease when performed by trained and skilled teams.	Systematic Review	Good
American College of Radiology, Appropriateness Criteria Chest Pain Suggestive of Acute Coronary Syndrome (ACR 2010a)	This guideline rates CCTA as 6 on a scale of 1-9[6] for patients with low to intermediate likelihood for CAD in the absence of cardiac enzyme or ischemic ST changes. It is a reasonable alternative to stress testing. It may be limited in patients with heart rates more than 65 beats/ minute, in patients with arrhythmias and in patients with calcium scores greater than 400-600 Agaston units.	Literature review, expert consensus	Fair
American College of Radiology, Appropriateness Criteria. Acute Chest Pain—Low Probability of CAD (ACR 2008)	This guideline rates CCTA as 7 on a scale of 1-9 for patients with acute chest pain and a low probability of CAD. CT of the chest may also detect other causes of chest pain.	Literature review, expert consensus	Fair
American College of Radiology, Appropriateness Criteria. Chronic chest pain with high probability of CAD (ACR, 2010b)	This guideline rates CCTA as 7 on a scale of 1-9 for patients with chronic chest pain and a high probability of CAD. It states that false negative studies may occur in the high risk group and negative studies may still require further diagnostic testing.	Literature review, expert consensus	Fair
ACCF/SCCT/ACR/AHA/ ASE/ASNC/SASCI/SCAI/ ACMR, 2010 Appropriate Use Criteria for CCTA (Taylor 2010)	This guideline lists multiple clinical scenarios and rates CCTA as inappropriate, appropriate or uncertain for patients with pretest probabilities of CAD of low intermediate and high for each clinical scenario. The resulting table is quite complex. Clinical scenarios and pretest probabilities for which CCTA is thought to be appropriate include: 1. Acute chest pain with suspicion of ACS and low to intermediate pretest probability of CAD with normal, uninterpretable or non-diagnostic ECG; 2. Non-acute symptoms possibly representing ischemic equivalent with normal, uninterpretable or non-diagnostic ECG; 3. Detection of CAD/Risk Assessment in asymptomatic patients without known CAD with low pretest probability and known family history of	Literature review, expert consensus	Fair

[6] The American College of Radiology Appropriateness Criteria gives a numerical score for the appropriateness of any radiology test for a given clinical setting. The scale used is 1 to 9 with 1 being the least appropriate and 9 being the most appropriate. The rating score is determined by expert consensus without published endpoints for individual scores. The American College of Radiology does not assign ranges that it considers “inappropriate” or “appropriate” but the relative scores for a series imaging tests for any clinical symptom does give some indication of relative appropriateness.

Recommending body, year published	Guideline(s)	Evidence base	Overall quality*
	premature CAD; 4. Detection of CAD/Risk Assessment in asymptomatic patients without known CAD with intermediate pretest probability and no family history of CAD; and 5. Continued symptoms with prior normal ECG stress test.		
ACCF/AHA Clinical Competence Statement on Vascular Imaging With Computed Tomography and Magnetic Resonance (Kramer 2007)	This guideline lists competence criteria for providers who perform CCTA. These include: 1. Knowledge of advances in CT hardware; 2. Competence in acquisition and interpretation techniques; 3. Knowledge of post-processing technique; and 4. Knowledge of contrast reactions and their treatment and techniques to reduce radiation dosage. The guideline recommends fellowship training, a minimum of 50 cases per year and 30 hours of CME training every 3 years.	Literature review, expert consensus	Fair

* The methodological quality of the guidelines was assessed using an instrument adapted from the instrument developed by the Appraisal of Guidelines Research and Evaluation (AGREE) Collaboration

Appendix F. Number of Asymptomatic Patients with True Positive and False Positive Results on CT Angiography Based on Prevalence of CAD in a Population*

Pre-test probability of CAD (prevalence if disease in the population)	**True positive test result** (number of patients with a positive test result who have CAD)	**False positive test result** (number of patients with a positive test result who do not have CAD)	**Post-test probability of CAD if the test result is "positive"** (positive predictive value)
1%	98	1485	6%
5%	490	1425	26%
10%	980	1350	42%
15%	1470	1275	54%
20%	1960	1200	62%
25%	2450	1125	68%
50%	4900	750	87%
75%	7350	375	95%

* Calculations were based on sensitivity of 98% and specificity of 85% from the meta-analysis by Ollendorf (2011) and 10,000 patients in a population. (See calculations below.) The pre-test probability for a patient may be estimated using clinical information and the Duke or Framingham risk scores.

Calculations for Appendix F

Obstructive Coronary Artery Disease (1% prevalence)

Diagnostic test result	Present n (%)	Absent n (%)	
Positive CCTA	98 (98%)	1485	1583 patients with a positive test result
Negative CCTA	2	8415 (85%)	8417 patients with a negative test result
Totals	100	9900	10,000 patients

Obstructive Coronary Artery Disease (5% prevalence)

Diagnostic test result	Present n (%)	Absent n (%)	
Positive CCTA	490 (98%)	1425	1915 patients with a positive test result
Negative CCTA	10	8075 (85%)	8085 patients with a negative test result
Totals	500	9500	10,000 patients

Obstructive Coronary Artery Disease (10% prevalence)

Diagnostic test result	Present n (%)	Absent n (%)	
Positive CCTA	980 (98%)	1350	
Negative CCTA	20	7650 (85%)	
Totals	1000	9000	10,000 patients

Obstructive Coronary Artery Disease (15% prevalence)

Diagnostic test result	Present n (%)	Absent n (%)	
Positive CCTA	1470 (98%)	1275	
Negative CCTA	30	7225 (85%)	
Totals	1500	8500	10,000 patients

Obstructive Coronary Artery Disease (20% prevalence)

Diagnostic test result	Present n (%)	Absent n (%)	
Positive CCTA	1960 (98%)	1200	
Negative CCTA	40	6800 (85%)	
Totals	2000	8000	10,000 patients

Obstructive Coronary Artery Disease (25% prevalence)

Diagnostic test result	Present n (%)	Absent n (%)	
Positive CCTA	2450 (98%)	1125	3575 patients with a positive test result
Negative CCTA	50	6375 (85%)	6425 patients with a negative test result

Totals	2500	7500	10,000 patients

Obstructive Coronary Artery Disease (50% prevalence)

Diagnostic test result	Present n (%)	Absent n (%)	
Positive CCTA	4900 (98%)	750	5650 patients with a positive test result
Negative CCTA	100	4250 (85%)	4350 patients with a negative test result
Totals	5000	5000	10,000 patients

Obstructive Coronary Artery Disease (75% prevalence)

Diagnostic test result	Present n (%)	Absent n (%)	
Positive CCTA	7350 (98%)	375	7725 patients with a positive test result
Negative CCTA	150	2125 (85%)	2275 patients with a negative test result
Totals	7500	2500	10,000 patients

Obstructive Coronary Artery Disease (90% prevalence)

Diagnostic test result	Present n (%)	Absent n (%)	
Positive CCTA	8820 (98%)	150	8970 patients with a positive test result
Negative CCTA	180	850 (85%)	1030 patients with a negative test result
Totals	9000	1000	10,000 patients

Appendix G. Analysis of Select Medicare, State and Private Payor CCTA Coverage Policies

LCD Number	State	Limits coverage to patients with low-intermediate pretest probability of CAD	Sets forth other patient criteria (e.g. obesity)	Covers for patient with known CAD	Covers Screening in asymptomatic patients	Requires 64 slice scanner	Sets forth physician competence expectations for performing procedure
L30047	**Alabama (Part A+B)**	No	No	Yes	No	Rec'd only	Yes
L24692	**Alaska Oregon Washington Minnesota (Part A)**	No	No	No	No	Yes	Yes
L23654	**Alaska Oregon Washington (Part B)**	No	No	No	No	Yes	Yes
L22038	**Arkansas (Part B)**	No	Yes	Yes	No	No	Yes
L24781	**Arkansas (Part A)**	No	Yes	Yes	No	No	Yes
L30288	**Minnesota (Part B) Missouri (Part A+B)**	No	No	No	No	Yes	Yes
L26751	**Oklahoma (Part A+B)**	No	Yes	Yes	No	No	Yes
L25907	**West Virginia (Part A)**	No	No	Yes	No	Yes	Yes
L22559	**West Virginia (Part B)**	Yes	Yes	No	No	Yes	Yes
AETNA	**-**	Yes (Low only)	Yes	No	No	Yes	No
CIGNA	**-**	Yes	No	No	No	Yes	No
MN DHS	**Minnesota (state policy)**	No	No	No	No	Yes	No
WA DSHS	**Washington (state policy)**	Yes	No	No	No	Yes	No

Appendix H. Select Medicare, State and Private Payor CCTA Coverage Policies

1. **L30047 – Radiology: Computed Tomography Angiography of the Heart and Coronary Vessels (FI and C)**
   State/Region: Alabama
   Contractor: Cahaba Government Benefit Administrators, LLC
   Contractor Type: MAC – Part A
   Effective Date: 05/04/2009 (original determination); 10/01/2010 (revisions)
   Indications
   As an alternative to invasive coronary angiography following a stress test that is equivocal or suspected to be inaccurate.

   Instead of myocardial perfusion imaging in the evaluation of coronary artery disease in those patients who have moderate pre-test probability of disease based on clinical risk factors and abnormal diagnostic studies, not symptoms alone.

   To evaluate the cause of symptoms in patients with known coronary artery disease.

   Assessment of suspected congenital anomalies of coronary circulation or great vessels.

   Assessment of coronary or pulmonary venous anatomy for the procedures described below:

   - CTA of the coronary veins is indicated when imaging of the coronary venous anatomy is necessary for biventricular pacemaker lead insertion.
   - CTA of the pulmonary veins is indicated when imaging of the pulmonary vasculature is necessary for pulmonary vein catheter ablation procedures for atrial fibrillation.

   Limitations
   Since the majority of the clinical research utilized a 64-slice CT scanner it is the recommended equipment. However, the intent of this LCD is not to monitor equipment utilization.

   The procedure must be performed under the direct supervision of and interpreted by a cardiologist or radiologist who meets the competency guidelines outlined by the published guidelines, ACCF/AHA Clinical Competence Statement on Cardiac Imaging with Computed Tomography and Magnetic Resonance, or American College of Radiology Clinical Statement on Noninvasive Cardiac Imaging.

   NOT COVERED:
   CPT 75571
   Using 71275 or 76497
   The procedure is not expected to be performed on an emergency department patient.

   Screening tests are defined as those tests done in the absence of signs, symptoms, or presence of disease. The use of these procedures (75572, 75573, 75574 for coronary CT angiography) in patients without signs, symptoms or presence of disease is considered to be screening by this Contractor.

2. **L24692 – Multidetector Computed Tomography of the Heart and Great Vessels (FI)**
State/Region: Alaska, Oregon, Washington, and Minnesota
Contractor: Noridian Administrative Services, LLC
Contractor Type: FI
Effective Date: 06/01/2007 (original determination); 10/01/2010 (revisions)
Note: Providers should seek information related to National Coverage Determinations (NCD) and other Centers for Medicare & Medicaid Services (CMS) instructions in CMS Manuals. This LCD only pertains to the contractor's discretionary coverage related to this service.

This medical policy consolidates and replaces all previous policies and publications on this subject by Noridian Administrative Services (NAS) and its predecessors for Medicare A.

Multidetector Computed Tomography (MDCT) with its advanced spatial and temporal resolution has opened up new possibilities in the imaging of the heart, coronary arteries, and major vessels of the chest. The MDCT technology uses thin (up to 1 mm) slices, 0.5 to 0.75 mm reconstructions, multiple simultaneous images (e.g. 16 or more "slices") and cardiac gating (often requiring beta blockers for ideal heart rate). There is significant post processing, depending on the number of slices per second, for image generation. Coronary artery images show a high correlation both with stenotic lesions noted on diagnostic cardiac catheterization and with atheromas on intracoronary ultrasound. Current evidence demonstrates that computed tomographic coronary angiography (CTCA) can reliably rule out the presence of significant coronary artery disease (CAD) in a patient with a low to intermediate probability of having CAD and can reliably achieve the high degree of diagnostic accuracy necessary to avoid conventional angiography in selected patients with a negative study (high predictive value of a negative study).

Indications
The **only** covered indications are as follow.

1. CTCA is covered for the evaluation of patients with acute chest pain presenting in an emergency room (or equivalent) when necessary to rapidly differentiate among reasonably probable aortic, pulmonary and/or coronary etiologies.

2. CTCA is covered as first-line testing for CAD in patients with low to intermediate risk factors* presenting in an emergency room with chest pain or other symptoms strongly suggestive of coronary disease (or in another site with an equivalent acute presentation). Enzymes and EKGs must be normal or borderline and the provider believes a negative CTCA will avoid invasive coronary angiography.

3. CTCA is covered for the exclusion of CAD following an equivocal or discordant or suspected inaccurate stress (or stress imaging) test in patients with low to intermediate risk factors the provider believes a negative CTCA will avoid invasive coronary angiography.

4. CTCA is covered for the evaluation of patients in sinus rhythm scheduled to undergo non-coronary (e.g., valvular) cardiovascular surgery who are unlikely to have coronary artery disease and/or significantly calcified coronary arteries and the provider believes a negative CTCA will avoid invasive coronary angiography.

5. CTCA or MDCT of the heart and great vessels is covered to assess surgical eligibility or for preoperative planning, in patients who have clinical findings strongly suggestive of a congenital anomaly of the coronary vessels or great vessels.

6. MDCT of the heart is covered for evaluation of pulmonary veins and atrium in patients with atrial fibrillation and/or flutter when evaluation avoids what would otherwise be a medically reasonable and necessary MRI in patients who are scheduled to undergo ablation therapy evaluation.

*For the purposes of this policy, “intermediate risk” is defined as TIMI < 4 and/or other equivalent accepted national standard, i.e., a standard describing the same or comparable level of risk.

Limitations

1. These tests are never covered for screening. Ultrafast CT scan of the heart (electron-beam tomography [**EBT**] or electron-beam computed tomography [**EBCT**]) is **not** a covered service.

2. The value of MDCT or CTCA for "risk stratification" in patients or scenarios not described in the Indications section of this LCD has not been sufficiently established and this use is non-covered.

3 Demonstration and/or quantification of the presence of coronary calcification in either asymptomatic or symptomatic patients with or without signs of atherosclerotic heart disease have not been shown to improve outcomes and is not covered. Until such time as there may be more evidence of medical necessity, Medicare will not pay for the quantitative evaluation of coronary calcium by MDCT, CTCA, EBCT or other technology.

4. At the initiation of any of these procedures, there must be an initial scout radiograph or other imaging assessment of cardiac calcification. The physician must make an assessment of the anatomic location, degree and intensity of calcification, and impact of calcification on the utility of test results. The CTCA will be subject to denial when post-pay review indicates that calcification of the coronary segment(s) in question, or other likely causes of significant artifact, will lead to invasive angiography to establish diagnosis.

5. Both false positive findings and number of non-evaluable segments increase proportionately with coronary calcification and may cause conversion to or result in invasive coronary angiography. Therefore, in patients with an overall Agatson calcium score of 600 or greater, documentation must indicate the medical necessity of proceeding with the examination, rather than converting to invasive arteriography". (For example, the physician might note the absence of significant calcification in the area of interest to support continuing with this study but rather proceed to conventional arteriography.

6. The selection of the test should be made within the context of other testing modalities and the resulting information should be essential to the management decision, not merely an additional layer of testing.

7. The administration of beta blockers and/or other drugs necessary for the study, any cardiograms, rhythm strips, IV drugs and other supplies and the monitoring of the patient during MDCT by a physician experienced in the use of cardiovascular drugs are included herein and are not separately payable services.

8. All studies must be ordered by a physician or a qualified non-physician practitioner.

9. A physician must be present for direct supervision during testing. (This requirement is presumed to be met in the hospital outpatient department.)

10. Coverage of this modality for coronary artery assessment is limited to devices that process thin, high resolution slices. Less resolution and slower rotation speeds result in a higher number of nonevaluable segments. Based on current literature, Medicare requires the multidetector scanner to either have *both* collimation of <u>0.625 mm or less, and a rotational speed of 375 msec or less OR, alternatively, at least 64 slice detector design</u>. Machines not meeting these requirements should not have studies submitted for payment.

11. Only one study, MDCTA of the heart and great vessels or chest CT, will be covered on a single date of service.

12.The MDCTA evaluation of both cardiac structure, morphology, and ventricular function in congenital heart disease (CPT 75573) remains non-covered by Medicare until such time as there is more evidence of the medical necessity for these procedures, including the specific clinical circumstances that describe both where and why these studies are needed for patient management]

13. Until such time as there is more evidence of the medical necessity for quantitative evaluation of coronary calcium, Medicare may not cover the procedure for coronary calcium scoring (75571).

Acceptable Levels of Competence for Performance and Interpretation
The acceptable levels of competence, as defined by the American College of Cardiology (ACC)/American Heart Association (AHA) Clinical Competence Statement on Cardiac Imaging with Computed Tomography and Magnetic Resonance (2005) and the American College of Radiology (ACR) Clinical Statement on Noninvasive Cardiac Imaging (2005), are outlined as follows:

"For the technical portion, a recommended level of competence is fulfilled when the image acquisition is obtained under all of the following conditions:

a. The service is performed by a radiologic technologist who is credentialed by a nationally recognized credentialing body (American Registry of Radiologic Technologists or equivalent) and meets state licensure requirements where applicable.

b. If intravenous beta blockers or nitrates are to be given prior to a CT coronary angiogram or calcium score, the test must be under the direct supervision of a certified registered nurse and physician (familiar with the administration of cardiac medications) who are available to respond to medical emergencies and it is strongly recommended that the certified register nurse and physician be ACLS certified.

c. When contrast studies are performed, the physician must provide direct supervision and the radiologic technologist or registered nurse administering the contrast must have appropriate training on the use and administration of contrast media."

For the professional portion, a recommended level of competence is fulfilled when the interpretation is performed by a physician meeting the following requirements:

"a. The physician has appropriate additional training in CT Coronary Angiography and cardiac CT imaging equivalent to the guidelines set forth by the ACC or ACR (for example: the ACCF/AHA Clinical Competence Statement on Cardiac Imaging with Computed Tomography and Magnetic Resonance (2005) and the ACR Clinical Statement on Noninvasive Cardiac Imaging (2005)), or

b. The physician has appropriate medical staff privileges to interpret CT Coronary Angiograms at a hospital that participates in the Medicare program, and is actively training in cardiac CT (as in paragraph a). A grace period of 24 months should be allowed to acquire the necessary training."

3. **L23654 – Multidetector Computed Tomography of the Heart and Great Vessels (Carrier)**
State/Region: Alaska, Oregon, Washington
Contractor: Noridian Administrative Services, LLC
Contractor Type: Carrier
Effective Date: 05/20/2007 (original determination); 10/01/2010 (revisions)

Coverage determination same as L24692.

4. **L22038 – Cardiac Computed Tomography (CCT) (Carrier)**
State/Region: Arkansas
Contractor: Pinnacle Business Solutions, Inc. - Arkansas
Contractor Type: Carrier
Effective Date: 02/15/2006 (original determination); 05/01/2010 (revisions)

A. Indications:
The Multi-detector Computed Tomography) (MDCT) (cardiovascular computed tomography (CCT)) may be employed for the following:

1. Emergency evaluation of acute chest pain syndrome for coronary etiology, including emergency evaluation, pulmonary embolism, aortic dissection, and coronary artery disease;

2. Cardiac evaluation of a patient with chest pain syndrome (e.g., anginal equivalent, angina) who is not a candidate for cardiac catheterization;

3. Management of a symptomatic patient with known coronary artery disease (e.g., post-stent, post CABG) when the results of the MDCT may guide the decision for repeat invasive intervention;

4. Assessment of suspected congenital anomalies of coronary circulation or great vessels; and

5. Assessment of coronary veins prior to biventricular pacing lead placement.

B. Limitation
1. Coverage of CT coronary angiography is limited to CT devices that process thin (up to 1 mm) slice, 0.5 to 0.75 mm reconstruction, and multiple simultaneous images.

2. The selection of the test should be made within the context of other testing modalities so that the resulting information facilitates the management decision, not merely adds a new layer of testing. Patient selection should be made with a high pretest probability of disease and must not be used for screening. This includes:

- Patients with irregular heart rhythm.
- Patients who have difficult breath hold, which of course will be much better with the 64 slice CT.
- Patients with extreme morbid obesity will be a limitation.
- Patients with serious intravenous iodinated contrast allergies.
- Patients with renal insufficiency for fear of contrast induced nephropathy.
- Heavily calcified coronary arteries would be a limiting factor for exact determination of diameter stenosis.
- Patients with acute coronary syndrome with chest pain and S-ST segment changes should go directly for invasive coronary angiogram.
- Radiation exposure should be considered as one of the limiting factors.

3. Electron Beam CT (EBCT) is not covered for use in coronary artery examination (Applicable to 75571 in which calcium scoring is done.)

4. The patient's treating physician or qualified non-physician practitioner must order the study.

5. Studies must be conducted under the direct supervision of a cardiologist and/or radiologist. IV beta blockers may be administered by a qualified non-physician practitioner as long as the direct supervision requirements are met.

6. All studies must be done by staff trained and accredited to do Computed Tomography. The supervising and interpreting physician must have appropriate additional training in CT Coronary Angiography and cardiac CT

imaging, equivalent to the guidelines set forth by the ACC (for cardiologists) and ACR (for radiologists), as found in the ACCF/AHA Clinical Competence Statement on Cardiac Imaging with Computed Tomography and Magnetic Resonance and the ACR Clinical Statement on Noninvasive Cardiac Imaging. A team approach for reading cardiac and non-cardiac computed tomography of the chest is suggested. This team-based approach assures that the beneficiary receives the newest in technology in the most competent hands. Once accreditation guidelines are in place, policy will be revised accordingly.

5. **L24781 – Cardiac Computed Tomography (CCT) (FI)**
   State/Region: Arkansas
   Contractor: Pinnacle Business Solutions, Inc. - Arkansas
   Contractor Type: FI
   Effective Date: 08/01/2006 (original determination); 05/01/2010 (revisions)

   Coverage determination same as L22038.

6. **L30288 – Computed Coronary Tomography Angiography (Carrier)**
State/Region: Minnesota Part B, Wisconsin, Missouri Part A & B
Contractor: Wisconsin Physicians Service Insurance Corporation
Contractor Type: Carrier
Effective Date: 08/16/2009 (original determination); 02/21/2011 (revisions)
Multi-slice or Multi-detector Computed Tomography (MDCT) with its advanced spatial and temporal resolution has opened up new possibilities in the imaging of the heart and major vessels of the chest, including the coronary arteries.

The MDCT technology requires thin (up to 1mm) slices, 0.5 to 0.75 mm reconstructions, multiple simultaneous images (e.g. 16, 32, 64 or more slices) and cardiac gating (often requiring beta blockers for ideal heart rate).

The current available body of evidence appears to demonstrate that coronary CTA (CCTA) can reliably rule out the presence of significant coronary artery disease (CAD) in patient with a low to intermediate probability of having CAD and can reliably achieve a high degree of diagnostic accuracy necessary to replace conventional angiography in selected situations

In some circumstances, CCTA may be proposed instead of, or in addition to, other noninvasive cardiac tests. This is particularly useful in the commonly encountered clinical scenario of patients having an equivocal stress myocardial perfusion test. The information from CCTA may be used to guide further diagnostic evaluation and/or appropriate therapy (e.g., revascularization versus medical management) and this may over the long term influence the morbidity from CAD.

It is expected that the levels of competence for both the technical and professional components of the procedure will be in compliance with those guidelines defined by the American College of Cardiology (ACC) American Heart Association (AHA) Clinical Competence Statement on Cardiac Imaging with Computed Tomography and Magnetic Resonance (2005) and the American College of Radiology (ACR) Clinical Statement on Noninvasive Cardiac Imaging (2005).

Indications
1. Coronary CTA used as an alternative to invasive angiography, following a stress test that is equivocal or suspected to be inaccurate.

Coronary CTA might be used as a triage tool as an alternative to invasive coronary angiography in select patients who have an equivocal or suspected inaccurate stress (or stress imaging) test. The rationale is that a noninvasive coronary anatomic test (CCTA) might permit a separate method of assessing the coronary arteries which is different from a stress test and limit the number of normal invasive coronary angiograms. It could also help avoid missing serious coronary disease in those suspected of having an inaccurate stress test result.

2. Coronary CTA for suspected congenital anomalies of the coronary circulation.

Coronary CTA is used to assess patients suspected of having a congenital coronary anomaly. The cross-sectional nature of this technique allows one to definitively determine both the presence and possible future harm that could result from the anomaly. It is often used after an anomaly has been suspected following a different test such as prior invasive coronary angiogram. A coronary CTA is used to decide if surgery is indicated and for surgical planning.

3. Coronary CTA for evaluation of acute chest pain in the emergency department (ED).

The rationale for the application of coronary CTA in this setting is to quickly triage patients in order to rule out coronary artery disease as a possible cause of symptoms. It is hoped that the application of coronary CTA in the emergency room would limit resource use in chest pain patients who do not have coronary artery disease. It is preferable that CCTA in the ED be ordered by a cardiologist.

4. CTA for the assessment of coronary or pulmonary venous anatomy

This application of CTA for the coronary and pulmonary veins is primarily for pre-surgical planning. Coronary venous anatomy can be useful for the cardiologist who needs to place a pacemaker lead in the lateral coronary vein in order to resynchronize cardiac contraction in patients with heart failure. This may be helpful to guide biventricular pacemaker placement.

Pulmonary vein anatomy can vary from patient to patient. Pulmonary vein catheter ablation can isolate electrical activity from the pulmonary veins and allow for the elimination of recurrent atrial fibrillation. The presence of a pulmonary venous anatomic map may help eliminate procedural complications and allow for the successful completion of the procedure.

Limitations

1. The test is never covered for screening, i.e., in the absence of signs, symptoms or disease.
2. The selection of the test should be made within the context of other testing modalities such as stress myocardial perfusion images or cardiac ultrasound result so that the resulting information facilitates the management decision, not merely adds a new layer of testing.
3. The test may be denied, on post-pay review, as not medically necessary when used for cardiac evaluation of a patient where there is a pre-test knowledge of sufficiently extensive calcification of the coronary segment in question that would diminish the interpretive value.
4. Coverage of this modality for coronary artery assessment is limited to devices that process thin, high resolution slices (1 mm or less). The multidetector scanner must have at least 64 slices per rotation capability.
5. The administration of beta blockers and the monitoring of the patient during CCTA by a physician experienced in the use of cardiovascular drugs are included and are not separately payable services.
6. All studies must be ordered by a physician or a qualified non-physician practitioner similar to any other medical testing such as the stress myocardial perfusion imaging or ultrasound evaluation.
7. For contrast enhanced examinations a physician must be present for direct supervision during testing similar to the stress myocardial perfusion imaging.
8. The electron beam tomography (EBT) technology or Ultrafast CT is not covered by this LCD for coronary artery examination.
9. Atrial fibrillation by itself is not an indication; atrial fibrillation with planned ablation therapy is allowed.

7. **L26751 – Cardiac Computed Tomography (CCT) (MAC – Part A)**
State/Region: Oklahoma Part A & B
Contractor: TrailBlazer Health Enterprises, LLC
Contractor Type: MAC – Part A
Effective Date: 03/31/2008 (original determination); 01/01/2011 (revisions)

Notice: It is not appropriate to bill Medicare for services that are not covered (as described by this entire LCD) as if they are covered.

Indications:
The Multi-Detector Computed Tomography (MDCT) (Cardiovascular Computed Tomography (CCT)) may be employed for the following:

- Emergency evaluation of acute chest pain syndrome for coronary etiology, including emergency evaluation, pulmonary embolism, aortic dissection and coronary artery disease.
- Cardiac evaluation of a patient with chest pain syndrome (e.g., anginal equivalent, angina) who is not a candidate for cardiac catheterization.
- Management of a symptomatic patient with known coronary artery disease (e.g., post-stent, post-CABG) when the results of the MDCT may guide the decision for repeat invasive intervention.
- Assessment of suspected congenital anomalies of coronary circulation or great vessels.
- Assessment of coronary veins prior to biventricular pacing lead placement.

Limitations:
Coverage of CT coronary angiography is limited to CT devices that process thin (up to 1 mm) slice, 0.5 to 0.75 mm reconstruction, and multiple simultaneous images.

The selection of the test should be made within the context of other testing modalities so that the resulting information facilitates the management decision, not merely adds a new layer of testing. Patient selection should be made with a high pretest probability of disease and must not be used for screening. This includes:

- Patients with irregular heart rhythm.
- Patients who have difficult breath hold, which of course will be much better with the 64-slice CT.
- Patients with extreme morbid obesity will be a limitation.
- Patients with serious intravenous iodinated contrast allergies.
- Patients with renal insufficiency for fear of contrast-induced nephropathy.
- Heavily calcified coronary arteries would be a limiting factor for exact determination of diameter stenosis.
- Patients with acute coronary syndrome with chest pain and S-ST segment changes should go directly for invasive coronary angiogram.
- Radiation exposure should be considered as one of the limiting factors.

Electron Beam CT (EBCT) is not covered for use in coronary artery examination.

The patient's treating physician or qualified non-physician practitioner must order the study.

Studies must be conducted under the direct supervision of a cardiologist and/or radiologist. IV beta blockers may be administered by a qualified non-physician practitioner as long as the direct supervision requirements are met.

All studies must be done by staff trained and accredited to perform computed tomography.

The supervising and interpreting physician must have appropriate additional training in CT coronary angiography and cardiac CT imaging, equivalent to the guidelines set forth by the ACC (for cardiologists) and ACR (for radiologists), as found in the ACCF/AHA Clinical Competence Statement on Cardiac Imaging with Computed Tomography and Magnetic Resonance and the ACR Clinical Statement on Non-Invasive Cardiac Imaging.

The CPT Category III Cardiac Computed Tomography Angiography (CCTA) codes include thorough review and reporting on all of the CT source images acquired. Per the ACR guidelines, non-cardiac structures imaged at the time of cardiac imaging must be reviewed and reported for pathology in addition to the cardiac structures. Medicare expects that when the CPT Category III CCTA codes are reported, all of the work described by the codes will have been performed. Although a physician may elect to have a separate physician interpret a portion of the images (e.g., non-cardiac structure images interpreted by a radiologist and cardiac structures interpreted by a cardiologist) only one professional component per study may be reported to Medicare regardless of the number of physicians contributing to the overall interpretation.

8. **L25907 - Cardiac Computed Tomography (CCT) and Coronary Computed Tomography Angiography (FI)**
State/Region: Virginia, West Virginia
Contractor: National Government Services, Inc.
Contractor Type: FI
Effective Date: 12/01/2007 (original determination); 01/01/2011 (revisions)

Abstract:
The multidetector helical computed tomography (MDCT) technology requires thin (up to 1 mm) slices, 0.5 to 0.75 mm reconstructions, multiple simultaneous images (e.g. 16, 32, 64 or more slices), and cardiac gating (often requiring beta blockers for ideal heart rate). There is significant post-processing, depending on the number of slices per second for image generation. For coronary artery imaging, the resulting images show a high correlation with stenotic lesions noted on diagnostic cardiac catheterization but more importantly, with atheromas on intracoronary ultrasound.

Current available body of evidence demonstrates that CCTA can reliably rule out the presence of significant coronary artery disease (CAD) in patients with a low to intermediate probability of having CAD and can reliably achieve a high degree of diagnostic accuracy and technical performance necessary to replace conventional angiography.

Indications:
Patient presenting with chest pain syndrome. CCTA may be used in lieu of an imaging stress test. The clinician must have a high degree of suspicion that CAD is high on the differential diagnosis of the symptoms.

To facilitate the management decision of a patient with an equivocal stress test. CCTA might be chosen in select patients who have an equivocal stress (or stress imaging) test. The rationale is that a noninvasive coronary anatomic test (CCTA) allows an alternate method of assessing the coronary arteries, which would limit the number of negative invasive coronary angiograms.

When the recurrence of symptoms in patients with known coronary artery disease may be related to progression/exacerbation of underlying disease. The use of CCTA in this setting would be to evaluate the extent of previously diagnosed coronary artery disease. Patients with known disease may have had remote invasive angiography and/or stress testing to evaluate prior events or symptoms. New or recurrent symptoms may relate to a change in the coronary anatomy that can be assessed with CCTA.

When patients with prior bypass surgery or intracoronary artery stent placement present with chest pain or dyspnea. Coronary bypass grafts are relatively well seen with CCTA. The rationale for CCTA would be to determine the patency and severity of possible graft stenoses that may be the source of chest pain. Patients with prior intracoronary stents often present with recurrent chest pain. The rationale for a CCTA as an alternative to invasive angiography is to rule out in-stent restenosis as the cause of symptoms. (Accurate assessment of in-stent restenosis may be limited by the artifact caused by the stent material itself and the quality of the scan and scanner).

Suspected congenital anomalies of the coronary circulation. CTA is used to assess patients suspected of having a congenital coronary anomaly. The cross-sectional nature of this technique allows one to determine accurately both the presence and possible future harm that could result from the anomaly. It is often used after an anomaly has been identified following a different test such as prior invasive coronary angiogram. A CCTA is used to decide if surgery is indicated and for surgical planning.

The assessment of coronary or pulmonary venous anatomy. This application of CTA for the coronary and pulmonary veins is primarily for pre-surgical planning. Coronary venous anatomy can be useful for the cardiologist who needs to place a pacemaker lead in the lateral coronary vein in order to resynchronize cardiac contraction in patients with heart failure. This may be helpful to guide biventricular pacemaker placement. Pulmonary vein anatomy can vary from patient to patient. Pulmonary vein catheter ablation can isolate electrical activity from the pulmonary veins and allow for the elimination of recurrent atrial fibrillation. The presence of a pulmonary venous

anatomic map may help eliminate procedural complications and allow for the successful completion of the procedure.

The patient undergoing non-coronary artery cardiac surgery. Certain patients who have non-coronary artery cardiac surgery (valve or ascending aortic surgery) may need a pre-operative invasive coronary angiogram. The surgical planning may also depend upon the exact location of the coronary arteries. The rationale for the use of CCTA in these patient subsets is to avoid potentially unnecessary invasive testing and still provide appropriate pre-surgical information.

The test may be medically necessary in patients presenting to the emergency room with complaints consistent with cardiac ischemia, but without diagnostic electrocardiography (ECG) or enzymes.

The test may be considered medically necessary in patients status post revascularization procedures who present with recurrent symptoms not clearly identifiable as ischemic.

Limitations:
The test is never covered for screening, i.e., in the absence of signs, symptoms or disease.

The test will be considered not medically necessary if the anticipated results are not expected to provide new, additional information to that already previously obtained from other tests (such as stress myocardial perfusion images or cardiac ultrasound). New or additional information should facilitate the management decision, not merely add a new layer of testing.

For dates of service prior to 01/01/2010, determination of cardiac ejection fraction (CPT code 0151T) should not be billed when previously determined by other techniques. CPT code 0151T is deleted effective 12/31/2009.

The test will be considered not medically necessary if it is anticipated that the patient would require invasive cardiac angiography for further diagnosis or for therapeutic intervention. (e.g., angina decubitus, unstable angina, Prinzmetal angina, etc.)

The test may be denied, on post-pay review, as not medically necessary when used for cardiac evaluation if there were pre-test knowledge of sufficiently extensive calcification of the suspect coronary segment that would diminish the interpretive value.

Effective 12/01/2009, coverage for evaluation of coronary artery or bypass graft stenosis, or for functional status (e.g., wall motion), is limited to multidetector scanners having at least 64 slices per rotation capability. This two year period (12/01/2007 - 12/01/2009 will allow for a phase-in of new technology.

The administration of beta blockers and the monitoring of the patient during MDCT/CCTA by a physician experienced in the use of cardiovascular drugs is included as part of the test and is not a separately payable service.

All studies must be ordered by the physician/qualified non-physician practitioner treating the patient and who will use the results of the test in the management of the patient.

The test must be performed under the direct supervision of a physician.

This LCD does not address electron beam tomography (EBT) technology or Ultrafast CT for coronary artery examination. There is no extension of coverage of EBT based on this policy.

Quantitative calcium scoring (CPT code 0144T for dates of service prior to 01/01/2010, and CPT 75571 on or after 01/01/2010) is not a covered service and will be denied as not medically necessary. Calcium scoring reported in isolation is considered a screening service. When performed in association with CT angiography, there is neither separate nor additional included reimbursement for the calcium scoring.

Acceptable Levels of Competence for Performance and Interpretation: Providers submitting claims for these tests must demonstrate proficiency and training in performing the tests according to the following standards:

The acceptable levels of competence, as defined by the American College of Cardiology (ACC)/American Heart Association (AHA) Clinical Competence Statement on Cardiac Imaging with Computed Tomography and Magnetic Resonance (2005) and the American College of Radiology (ACR) Clinical Statement on Noninvasive Cardiac Imaging (2005), are outlined as follows:

For the technical portion, a recommended level of competence is fulfilled when the image acquisition is obtained under all of the following conditions:

The service is performed by a radiology technologist who is credentialed by a nationally recognized credentialing body (American Registry of Radiologic Technologists or equivalent) and meets state licensure requirements where applicable.

If intravenous beta blockers or nitrates are to be given prior to a CT coronary angiogram, the test must be under the direct supervision of a certified registered nurse and physician (familiar with the administration of cardiac medications) who are available to respond to medical emergencies and it is strongly recommended that the certified register nurse and physician be ACLS certified.

When contrast studies are performed, the physician must provide direct supervision and the radiologic technologist or registered nurse administering the contrast must have appropriate training on the use and administration of contrast media.

For the professional portion, a recommended level of competence is fulfilled when the interpretation is performed by a physician meeting the following requirements:

The physician has appropriate additional training in coronary CTA and cardiac CT imaging equivalent to the guidelines set forth by the ACC or ACR (for example: the ACCF/AHA Clinical Competence Statement on Cardiac Imaging with Computed Tomography and Magnetic Resonance (2005) and the ACR Clinical Statement on Noninvasive Cardiac Imaging (2005)), or

The physician has appropriate medical staff privileges to interpret CT coronary angiograms at a hospital that participates in the Medicare program, and is actively training in cardiac CT (as in paragraph a). A grace period of 24 months will be allowed to acquire the necessary training.

9. **L22559 – Cardiac Computed Tomography & Angiography (CCTA) (Carrier)**
State/Region: West Virginia
Contractor: Palmetto GBA
Contractor Type: Carrier
Effective Date: 10/01/2006 (original determination); 10/01/2010 (revisions)
Indications and Limitations of Coverage
Cardiac computed tomographic angiography (CCTA), also known as computed tomography of the heart and coronary arteries, or multidetector computed cardiac tomography (MDCT) is considered reasonable and necessary for the evaluation of suspected symptomatic coronary artery disease (CAD) and for the detection of structural and morphologic intra- and extra-cardiac conditions.

Use of a CCTA is expected to avoid diagnostic cardiac catheterization. If high pre-test probability of CAD exists, Palmetto expects the patient to undergo invasive coronary angiography with appropriate percutaneous coronary intervention.

To establish CCTA medical necessity, your case must meet at least one indication in the following two categories:

Symptomatic Coronary Artery Disease (CAD)
1. Evaluation of Acute Chest Pain, unexplained dyspnea or symptoms suggesting angina pectoris (such as jaw pain) when there is:
   - Intermediate pre-test probability of CAD*, **and**
   - No EKG changes to suggest acute myocardial injury or ischemia, **and**
   - Normal initial cardiac markers.

2. Evaluation of Chest Pain Syndrome, when there is:
   - Intermediate pre-test probability of CAD*, **and**
   - Uninterpretable EKG** **or** patient is unable to exercise, **or**
   - Uninterpretable **or** equivocal stress test (exercise, perfusion **or** stress echo)

*Intermediate pretest probability of CAD by age, gender and symptoms is between 10 and 90% as referenced in the ACCF/ACR 2006 Appropriateness Criteria for Cardiac Computed Tomography and Cardiac Magnetic Resonance Imaging.

** Uninterpretable EKG refers to EKGs with resting ST segment depression greater than or equal to 0.10mV, complete left bundle branch block, pre-excitation, or paced rhythm.

3. Evaluation of intracardiac structures for suspected coronary anomalies.

Suspected Cardiac Structural/Morphologic Anomalies
1. Detection of intracardiac and extracardiac structures in:
   - Evaluation of cardiac mass (suspected tumor or thrombus) **or**
   - Evaluation of pericardial conditions (mass, constrictive pericarditis, **or** complications of cardiac surgery), **or**
   - Patients with technically limited images from echocardiogram, MRI **or** TEE.

2. Detection of morphologic intracardiac and extracardiac structures for:
   - Evaluation of pulmonary vein anatomy prior to invasive radiofrequency ablation for atrial fibrillation. While data is limited for 3D reconstruction of the left atrium for ablations, there is broad consensus among cardiologists that these images, which are integrated and used in real-time in the procedure room to shorten procedure time, improve therapeutic success and enhance patient safety, **or**
   - Non-invasive coronary vein mapping prior to placement of biventricular pacemaker, **or**
   - Non-invasive coronary arterial mapping, including internal mammary artery, prior to repeat cardiac

surgical revascularization, **or**
- Detection of complex congenital heart disease including anomalies of coronary circulation, great vessels, and cardiac chamber and valves, **or**
- Evaluation of coronary arteries in patients with new onset heart failure to assess etiology.

Limitations:

1. Coverage of CCTA is limited to CT devices that process thin, high resolution slices. Decreased resolution and slower rotation speeds result in a higher number of non-evaluable segments. At the current time, Medicare requires the multidetector scanner to have collimation of 0.625 mm or less, and a rotational speed of 375 msec or less, OR to have at least 64 slice detector design. Do not submit studies from scanners that do not meet these requirements.

2. Medicare does **not** cover a screening CCTA for asymptomatic patients, risk stratification or quantitative evaluation of coronary calcium.

Ultrafast CT scan of the heart (electron-beam tomography [EBT] or electron-beam computed tomography [EBCT]) is **not** a covered service.

3. Simultaneous exclusion of obstructive CAD, pulmonary embolism, and aortic dissection ("triple rule-out") in the emergency department is **not** covered. In order to optimize imaging of the RCA, contrast must be cleared from the right sided chambers during acquisition, a process that leads to suboptimal contrast timing in the pulmonary arteries. Simultaneous rule-out of aortic pathology (at the low pitch needed to properly image the coronaries) mandates thicker slices in order to capture the total volume required in a reasonable breath hold. The increased slice thickness degrades coronary image quality.

4. For CCTA patients must be able to lie still, follow breathing instructions, take nitroglycerine for coronary dilatation and take a beta-blocker or calcium blocker to achieve heart rates less than 70 BPM.

5. Prior to the initiation of a CCTA, there must be an imaging assessment of coronary calcification (calcium scoring). The physician must make an assessment of the anatomic location, degree and intensity of calcification and impact of calcification on the utility of the test results. CCTAs performed on patients with elevated quantitative calcium scores that preclude accurate assessment of coronary anatomy are **not** covered by Medicare.

Palmetto GBA expects that CCTA is performed under the direct supervision of a physician with appropriate training in CT coronary angiography and cardiac CT imaging equivalent to guidelines set forth by the ACC or ACR (Circulation 2005:112:598-617/ J Am Coll Cardiol. 2005:46:383-402)

10. **Aetna Policy** – Clinical Policy Bulletin: Cardiac CT, Coronary CT Angiography and Calcium Scoring
State/Region: n/a
Contractor: n/a
Contractor Type: n/a
Effective Date: 04/09/1998 (original determination); 12/07/2010 (revisions)

1. Aetna considers cardiac computed tomography (CT) angiography of the coronary arteries using 64 slices or greater medically necessary for the following indications:

    A. Rule out significant coronary stenosis in persons with a low or very low pre-test probability of coronary artery disease by Framingham risk scoring or by American College of Cardiology criteria (see appendix), with any of the following indications:

        a. Evaluation of persons with chest pain who cannot perform or have contraindications to exercise and pharmacologic stress testing (see appendix); *or*
        b. Evaluation of persons with a positive (i.e., greater than or equal to 1 mm ST segment depression) exercise stress test; *or*
        c. Evaluation of persons with chest pain presenting to the emergency department when an imaging stress test or coronary angiography are being deferred as the initial imaging study.

    B. Evaluation of asymptomatic persons at low pretest probability of coronary heart disease by Framingham risk scoring (see appendix) who have an equivocal exercise or pharmacological stress test. Note: Current guidelines from the American Heart Association recommend against routine stress testing for screening asymptomatic adults.
    C. Preoperative assessment of persons scheduled to undergo 'high-risk" noncardiac surgery, where an imaging stress test or invasive coronary angiography is being deferred unless absolutely necessary. The American College of Cardiology defines high-risk surgery as emergent operations, especially in the elderly, aortic and other major vascular surgeries, peripheral vascular surgeries, and anticipated prolonged surgical procedures with large fluid shifts and/or blood loss involving the abdomen and thorax.
    D. Preoperative assessment for planned noncoronary cardiac surgeries including valvular heart disease, congenital heart disease, and pericardial disease, in lieu of cardiac catheterization as the initial imaging study.
    E. Detection and delineation of suspected coronary anomalies in young persons (less than 30 years of age) with suggestive symptoms (e.g., angina, syncope, arrhythmia, and exertional dyspnea without other known etiology of these symptoms in children and adults; dyspnea, tachypnea, wheezing, periods of pallor, irritability (episodic crying), diaphoresis, poor feeding and failure to thrive in infants).

2. Aetna considers CT angiography of cardiac morphology for pulmonary vein mapping medically necessary for the following indications:

    A. Evaluation of persons needing biventricular pacemakers to accurately identify the coronary veins for lead placement.
    B. Evaluation of the pulmonary veins in persons undergoing pulmonary vein isolation procedures for atrial fibrillation (pre- and post-ablation procedure).

3. Aetna considers cardiac CT for evaluating cardiac structure and morphology in congenital heart disease medically necessary for the following indications:

    A. Anomalous pulmonary venous drainage;
    B. Evaluation of other complex congenital heart diseases;
    C. Evaluation of sinus venosum atrial-septal defect;
    D. Kawasaki's disease;

E. Person scheduled or being evaluated for surgical repair of tetralogy of Fallot or other congenital heart disease;
F. Pulmonary outflow tract obstruction;
G. Suspected or known Marfan's syndrome.

4. Aetna considers cardiac CT angiography experimental and investigational for persons with any of the following contraindications to the procedure:

   A. Body mass index (BMI) greater than 40.
   B. Inability to image at desired heart rate (under 80 beats per minute), despite beta blocker administration.
   C. Person with allergy or intolerance to iodinated contrast material
   D. Persons in atrial fibrillation or with other significant arrhythmia.
   E. Persons with extensive coronary calcification by plain film or with prior Angston score greater than 1700.

5. Aetna considers coronary CT angiography experimental and investigational for screening of asymptomatic persons, evaluation of persons at intermediate or high pretest probability of coronary artery disease, evaluation of stent occlusion or in-stent restenosis, evaluation of persons with an equivocal PET rubidium study, and for all other indications.

   Aetna considers cardiac CT angiography using less than 64-slice scanners experimental and investigational.

6. Aetna considers calcium scoring medically necessary for diagnostic cardiac CT angiography to assess whether an adequate image of the coronary arteries can be obtained.

   Aetna considers calcium scoring (e.g., with ultrafast (electron beam) CT, spiral (helical) CT, and multi-slice CT) experimental and investigational for all other indications because the definitive value of calcium scoring for assessing coronary heart disease risk has not been established in the peer-reviewed published medical literature.

11. **CIGNA – Medical Coverage Policy: Computed Tomography Angiography (CTA)**
State/Region: all plans administered by CIGNA Companies, including plans administered by Great-West Healthcare
Contractor: n/a
Contractor Type: n/a
Effective Date: 08/15/2010 (original determination); 08/15/2011 (next review date)

CIGNA covers 64-slice or greater multidetector-row computed tomography angiography (CTA) as medically necessary as an adjunct to other testing for ANY of the following indications:

- evaluation of chest pain in an individual with a very low, low, or intermediate pre-test probability of coronary artery disease[1] (CAD) when the individual cannot perform or has a contraindication to exercise and chemical stress testing (i.e. exercise treadmill stress test, stress echo, and nuclear stress test [i.e., myocardial perfusion imaging])
- exclusion of CAD in an individual with a low or very low pre-test probability of CAD when recent stress test results (i.e., exercise treadmill, stress echo, or nuclear stress test [i.e., myocardial perfusion imaging]) are uninterpretable, equivocal, or there is a suspicion that the results are falsely positive
- exclusion of CAD in an individual with an intermediate pre-test probability of CAD when recent stress test results (i.e., exercise treadmill, stress echo, or nuclear stress test [i.e., myocardial perfusion imaging]) are uninterpretable or equivocal, AND CTA will be performed in lieu of an angiography.
- exclusion of CAD in a symptomatic individual (e.g., acute chest pain in an emergency department setting), and the individual has an intermediate pre-test probability of CAD, and there are no changes noted on the ECG and serial enzymes are negative
- evaluation of suspected or known coronary artery anomalies associated with congenital conditions
- for morphologic evaluation of the coronary arteries in an individual with dilated cardiomyopathy or new onset heart failure, when ischemia is the suspected etiology and cardiac catheterization and/or nuclear stress test (i.e., myocardial perfusion imaging) have not been performed
- pre-operative assessment of coronary arteries in an individual undergoing repair of aortic dissection, aortic aneurysm repair or valvular surgery AND CTA will be performed in lieu of an angiography
- post-coronary artery bypass grafting (CABG) when BOTH of the following criteria are met:
    - repeat intervention is being considered
    - recent coronary angiography has been completed but additional information is needed before a treatment decision can be made

CIGNA does not cover multidetector-row computed tomography angiography (CTA) for any other indication, including but not limited to those listed below, because it is considered experimental, investigational or unproven:

- evaluation of chest pain in an intermediate or high pre-test probability of CAD individual when recent stress test result (i.e., exercise treadmill, stress echo, or nuclear stress test [i.e., myocardial perfusion imaging]) are either clearly positive or unequivocally negative
- screening for CAD in an asymptomatic individual
- post-revascularization procedure (e.g., percutaneous coronary intervention, coronary artery bypass grafting surgery), including evaluation of bypass grafts, coronary anatomy or evaluation for in-stent restenosis except when an individual is post-coronary artery bypass grafting (CABG), repeat intervention is being considered but additional information is required following completion of recent coronary angiography.

[1] See Appendix A for definition of Pre-test Probability of CAD

12. **Minnesota DHS – Authorization Criteria: Computed Tomography of Coronary Arteries**
State/Region: Minnesota
Contractor: n/a
Contractor Type: n/a
Effective Date: 08/01/2010
MHCP will cover CT of cardiac coronary arteries when following are met.

Coverage Indications:
1. Coronary CTA requested as an alternative to invasive angiography, following a stress test that is equivocal or suspected to be inaccurate.
2. Coronary CTA for suspected congenital anomalies of the coronary circulation.
3. Coronary CTA for evaluation of acute chest pain in the emergency department (ED).
4. CTA requested for the assessment of coronary or pulmonary venous anatomy for pre-surgical planning

Contraindication/noncoverage:
1. The test is never covered for screening, i.e., in the absence of signs, symptoms or disease.
2. The selection of the test should be made within the context of other testing modalities such as stress myocardial perfusion images or cardiac ultrasound result so that the resulting information facilitates the management decision, not merely adds a new layer of testing.
3. Coverage of this modality for coronary artery assessment is limited to devices that process thin, high resolution slices (1 mm or less). The multidetector scanner must have at least 64 slices per rotation capability.
4. The administration of beta blockers and the monitoring of the patient during CCTA by a physician experienced in the use of cardiovascular drugs are included and are not separately payable services.
5. All studies must be ordered by a physician or a qualified non-physician practitioner similar to any other medical testing such as the stress myocardial perfusion imaging or ultrasound evaluation.
6. For contrast enhanced examinations a physician must be present for direct supervision during testing similar to the stress myocardial perfusion imaging.
7. The electron beam tomography (EBT) technology or Ultrafast CT is not covered for coronary artery examination.
8. Atrial fibrillation by itself is not an indication; atrial fibrillation with planned ablation therapy is allowed

13. **Washington State Health Care Authority – Health Technology Clinical Committee Findings and Coverage Decisions: Coronary Computed Tomography Angiography**
State/Region: Washington
Contractor: n/a
Contractor Type: n/a
Effective Date: 05/08/2009 (original determination)

HTCC Coverage Determination
Coronary Computed Tomographic Angiography (CCTA) is **covered benefits with conditions** consistent with the criteria identified in the reimbursement determination**.**

HTCC Reimbursement Determination
Limitations of Coverage

1) Patients with low to intermediate risk of coronary artery disease;

2) For investigation of acute chest pain in an emergency department or hospital setting; and

3) Using Computed Tomography machines with 64-slice or better capability.

Non-Covered Indicators

- Patients who are asymptomatic or at high risk of coronary artery disease;
- CCTA used for coronary artery disease investigation outside of the emergency department or hospital setting; and
- CT scanners that use lower than 64- slice technology.

Appendix I. Relevant Codes

CODES	DESCRIPTION
ICD 9	
786.5	Chest pain
786.50	Chest pain unspecified
786.59	Discomfort, pressure, tightness in chest
413	Angina pectoris
414.0	Coronary atherosclerosis
CPT	
75574	Coronary computed tomographic angiography (CCTA)
78451-4	Nuclear perfusion imaging
93350	Stress echocardiography
93015	Stress ECG
93454	Cather coronary angiography (ICA)

References

Abidov, A., Callagher, M.J., Cinnaiyan, K.M., Mehta, L.S., Wegner, J.H., & Raff, G.L. (2009). Clinical effectiveness of coronary computed tomographic angiography in the triage of patients to cardiac catheterization and revascularization after inconclusive stress testing: Results of a 2-year prospective trial. *Journal of Nuclear Cardiology*, *16*, 701-13.

AGREE Next Steps Consortium. (2009). *Appraisal of guidelines for research and evaluation II: Instrument*. Retrieved April 5, 2011, from http://www.agreetrust.org/about-agree/introduction0/

American College of Radiology. (2008). *ACR Appropriateness Criteria. Acute chest pain — Low probability of coronary artery disease*. Reston, VA: American College of Radiology. Retrieved February 22, 2011, from http://www.acr.org/SecondaryMainMenu Categories/quality_safety/app_criteria/pdf/ExpertPanelonCardiovascularImaging/AcuteChestPainNoECGorEnzymeEvidenceofMyocardialIschemiaInfarctionDoc1.aspx

American College of Radiology. (2010a). *ACR Appropriateness Criteria. Chest pain, suggestive of acute coronary syndrome.* Reston, VA: American College of Radiology. Retrieved February 22, 2011, from http://www.acr.org/SecondaryMainMenuCategories/quality_safety/app_criteria/pdf/ExpertPanelonCardiovascularImaging/Chest-Pain-Suggestive-of-Acute-Coronary-Syndrome.aspx

American College of Radiology. (2010b). *ACR Appropriateness Criteria. Chronic chest pain — High probability of coronary artery disease*. Reston, VA: American College of Radiology. Retrieved February 22, 2011, from http://www.acr.org/SecondaryMainMenuCategories/quality_safety/app_criteria/pdf/ExpertPanelonCardiovascularImaging/ChronicChestPainNoEvidenceofMyocardialIschemiaInfarctionUpdateinProgressDoc7.aspx

Bayliss, J. (2002). *Duke clinical score: Prediction of coronary heart disease in a patient with chest pain*. Retrieved April 5, 2011, from http://www.zunis.org/Duke%20Chest%20Pain%20-%20CAD%20Predictor.htm

Center for Disease Control and Prevention. (2010). *Heart disease*. Retrieved February 15, 2011, from http://www.cdc.gov/heartdisease/facts.htm

Centers for Medicare and Medicaid Services. (2008). *Decision memo for computed tomographic angiography (CAG-00385N).* Retrieved March 31, 2011, from http://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=206&ver=17&NCDId=176&ncdver=2&NcaName=Computed+Tomographic+Angiography&IsPopup=y&bc=AAAAAAAAEAAA&

Danciu, S.C., Herrera, C.J., Stecy, P.J., et al. (2007). Usefulness of multislice computed tomographic coronary angiography to identify patients with abnormal myocardial perfusion stress in whom diagnostic catheterization may be safely avoided. *American Journal of Cardiology, 100*(11), 1605–8.

Diamond, G.A., & Forrester, J.S. (1979). Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease. *New England Journal of Medicine, 300*(24), 1350-1358.

Gibbons, R.J. (1999). ACC/AHA/ACP-ASIM chronic stable angina guidelines. *Journal of the American College of Cardiology, 33,* 2092–197.

Goldstein, J.A., Gallagher, M.J., O'Neill, W.W., Ross, M.A., O'Neil, B.J., & Raff, G.L. (2007). A randomized controlled trial of multi-slice coronary computed tomography for evaluation of acute chest pain. *Journal of the American College of Cardiology, 49*, 863-871.

Guyatt, G.H., Oxman, A.D., Vist, G.E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., et al. (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *British Medical Journal, 336*(7650), 924-926.

Hadamitzky, M., Freissmuth, B., Meyer, T., Hein, F., Kastrati, A., Martinoff, S., et al. (2009). Prognostic value of coronary computed tomographic angiography for prediction of cardiac events in patients with suspected coronary artery disease. *Journal of the American College of Cardiology, 2*, 404-411.

Halpern, E.J., Savage, M.P., Fischman, D.L., & Levin, D.C. (2010). Cost-effectiveness of coronary CT angiography in evaluation of patients without symptoms who have positive stress test results. *American Journal of Roentgenology, 194*(5), 1257-1262.

Hamirani, Y.S., Isma'eel, H., Larijani, V., Drury, P., Lim, W., Bevinal, M., et al. (2010). The diagnostic accuracy of 64-detector cardiac computed tomography compared with stress nuclear imaging in patients undergoing invasive cardiac catheterization. *Journal of Computed Assisted Tomography, 34*(5), 645-651.

Hay, C.S.M., Morse, R.J., Morgan-Hughes, G.J., et al. (2009). Prognostic value of coronary multidetector CT angiography in patients with an intermediate probability of significant coronary heart disease. *British Journal of Radiology*. (Epub ahead of print).

Hess, E.P., Agarwal, D., Chandra, S., Murad, M.H., Erwin, P.J., Hollander, J.E., et al. 2010). Diagnostic accuracy of the TIMI risk score in patients with chest pain in the emergency department: A meta-analysis. *Canadian Medical Association Journal*, *182*(10), 1039-1044.

Hoffmann, U., Bamberg, F., Chae, C.U., Nichols, J.H., Rogers, I.S., Seneviratne, S.K., et al. (2009). Coronary computed tomography angiography for early triage of patients with acute chest pain. The ROMICAT (Rule Out Myocardial Infarction using Computer Assisted Tomography) trial. *Journal of the American College of Cardiology, 53*(18), 1642–50.

Hollander, J.E., Litt, H.I., Chase, M., et al. (2007). Computed tomography coronary angiography for rapid disposition of low-risk emergency department patients with chest pain syndromes. *Academic Emergency Medicine, 14*(2), 112-6.

Institute for Clinical and Economic Review (ICER). (2008). *Coronary computed tomographic angiography for detection of coronary artery disease*. Boston, MA: ICER. Retrieved

February 16, 2011, from http://www.hta.hca.wa.gov/documents/ccta_appraisal_101708_final.pdf

Johnson, T.R., Nikolaou, K., Busch, S., et al. (2007). Diagnostic accuracy of dual source computed tomography in the diagnosis of coronary artery disease. *Investigative Radiology, 42,* 684–91.

Karlsberg, R.P., Budoff, M.J., Thompson, L.E.J., Friedman, J.D., & Berman, J.S. (2010). Reduction in downstream test utilization following introduction of coronary computed tomography in a cardiology practice. *International Journal of Cardiovascular Imaging*, *26*, 359-366.

Kramer, C.M., Budoff, M.J., Fayad, Z.A., Ferrari, V.A., Goldman, C., Lesser, J.R., et al. (2007). ACCF/AHA clinical competence statement on vascular imaging with computed tomography and magnetic resonance. *Circulation, 116*(11), 1318-1335.

Ladapo, J.A., Jaffer, F.A., Hoffmann, U., Thomson, C.C., Bamberg, F., Dec, W., et al. (2009). Clinical outcomes and cost-effectiveness of coronary computed tomography angiography in the evaluation of patients with chest pain. *Journal of the American College of Cardiology, 54*(25), 2409-22.

Lloyd-Jones, D., Adams, R., Carnethon, M., De Simone, G., Ferguson, B., Flegal, K., et al. (2009). AHA statistical update: Heart disease and stroke statistics – 2009 update. A report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation, 122*, e1-e161.

May, J.M., Shuman, W.P., Strote, J.N., Branch, K.R., Mitsuori, L.M., Lockhart, D.W., et al. (2009). Low-risk patients with chest pain in the emergency department: negative 64-MDCT coronary angiography may reduce length of stay and hospital charges. *American Journal of Roentgenography*, *193*, 150-154.

Mayo Clinic. (2011a). *Coronary angiogram*. Retrieved February 15, 2011, from http://www.mayoclinic.com/health/coronary-angiogram/MY00541

Mayo Clinic. (2011b). *CT angiogram*. Retrieved February 15, 2011, from http://www.mayoclinic.com/health/ct-angiogram/my00670

National Institute for Health and Clinical Excellence (NICE). (2009). *The guidelines manual.* London: NICE. Retrieved October 4, 2010, from http://www.nice.org.uk/media/5F2/44/The_guidelines_manual_2009_-_All_chapters.pdf

Ollendorf, D.A., Kuba, M., & Pearson, S.D. (2010). The diagnostic performance of multi-slice coronary computed tomographic angiography: A systematic review. *Journal of General Internal Medicine, 26*(3), 307-316.

Panju, A.A., Hemmelgarn, B.R., Guyatt, G.H., & Simel, D.L. (1998). The rational clinical examination: Is this patient having a myocardial infarction. *Journal of the American Medical Association, 280*(14), 1256-1263.

Pryor, D.B., Harrell, F.E. Jr., Lee, K.L., Califf, R.M., & Rosati, R.A. (1983). Estimating the likelihood of significant coronary artery disease. *American Journal of Medicine, 75*, 771-80.

Pryor, D.B., Shaw, L., McCants, C.B., Lee, K.L., Mark, D.B., Harrell, F.E. Jr., et al. (1993). Value of the history and physical in identifying patients at increased risk for CAD. *Annals of Internal Medicine, 118*, 81-90.

Pundziute, G., Schuijf, J.D., Jukema, J.W., et al. (2007). Prognostic value of multislice computed tomography coronary angiography in patients with known or suspected coronary artery disease. *Journal of the American College of Cardiology, 49*(1), 62-70.

Rubinshtein, R., Halon, D.A., Gaspar, T., et al. (2007). Usefulness of 64-slice cardiac compute tomographic angiography for diagnosing acute coronary syndromes and predicting clinical outcome in emergency department patients with chest pain of uncertain origin. *Circulation, 115*(13), 1762-8.

Savino, G., Herzog, C., Costello, P., et al. (2006). 64 slice cardiovascular CT in the emergency department: Concepts and first experiences. La *Radiologica Medica, 111*(4), 481-96.

Scottish Intercollegiate Guidelines Network (SIGN). (2007). *Management of stable angina.* Edinburgh: SIGN. Retrieved February 16, 2011, from http://www.sign.ac.uk/pdf/sign96.pdf

Scottish Intercollegiate Guidelines Network. (2009). *Critical Appraisal: Notes and Checklists***.** Edinburgh: SIGN. Retrieved November 15, 2010, from http://www.sign.ac.uk/methodology/checklists.html

Schuetz, G.M., Zacharopoulou, N.M., Schlattmann, P., & Dewey, M. (2010). Meta-analysis: Noninvasive coronary angiography using computed tomography versus magnetic resonance imaging. *Annals of Internal Medicine, 152*, 167-177.

Taylor, A.J., Cerqueira, M., Hodgson, J.M., Mark, D., Min, J., O'Gara, P., et al. (2010). ACCF/SCCT/ACR/AHA/ASE/ASNC/NASCI/SCAI/SCMR 2010 appropriate use criteria for cardiac computed tomography. *Journal of the American College of Cardiology*, *56*, 1864-1894.

Than, M., Cullen, L., Read, C.M., et al. (2011). A 2-h diagnostic protocol to assess patients with chest pain symptoms in the Asia-Pacific region (ASPECT): A prospective observational validation study. *Lancet, 377*, 1077-84.

Wagdi, P., & Alkadhi, H. (2010). The impact of cardiac CT on the appropriate utilization of catheter coronary angiography. *International Journal of Cardiovascular Imaging, 26*, 333-344.

Weustink, A.C., Mollet, N.R., Neefjes, L.A., Meijboom, W.B., Galema, T.W., van Mieghem, C.A., et al. (2010). Diagnostic accuracy and clinical utility of noninvasive testing for coronary artery disease. *Annals of Internal Medicine*, *152*, 630-639.

INSTITUTE FOR CLINICAL AND ECONOMIC REVIEW

FINAL APPRAISAL DOCUMENT

CORONARY COMPUTED TOMOGRAPHIC ANGIOGRAPHY FOR DETECTION OF CORONARY ARTERY DISEASE

January 9, 2009

Senior Staff	
Daniel A. Ollendorf, MPH, ARM	Chief Review Officer
Alexander Göhler, MD, PhD, MSc, MPH	Lead Decision Scientist
Steven D. Pearson, MD, MSc	President, ICER
Associate Staff	
Michelle Kuba, MPH	Sr. Technology Analyst
Marie Jaeger, B.S.	Asst. Decision Scientist

© 2009, Institute for Clinical and Economic Review

CONTENTS

About ICER 3

Acknowledgments 4

Executive Summary 5

Evidence Review Group Deliberation 17

ICER Integrated Evidence Rating 25

Evidence Review Group Members 27

Appraisal Overview 30

Background 33

- **The Technology and its Comparators** 34
- **Clinical Guidelines & Competency Standards** 37
- **Medicare and Representative Private Insurer Coverage Policies** 39
- **Previous Systematic Reviews/Tech Assessments** 41
- **Ongoing Clinical Studies** 43

The Evidence 45

- **Systematic Literature Review** 45
- **Data Analyses** 47
- **Results** 49
- **Summary** 63

Decision Analytic/Economic Models 66

- **ED Model** 67
- **Outpatient Model** 75

Recommendations for Future Research 91

References 93

Systematic Review Tables 105

Appendices 112

ABOUT ICER

The Institute for Clinical and Economic Review (ICER) provides independent evaluation of the clinical effectiveness and comparative value of new and emerging technologies. ICER is based at the Massachusetts General Hospital's Institute for Technology Assessment (ITA), an affiliate of Harvard Medical School. ICER develops its assessments in collaboration with faculty and staff from the ITA and Harvard Medical School as well as with researchers and clinical experts from around the country. All ICER assessments are performed in conjunction with an external Evidence Review Group comprised of clinical and policy experts who serve a longitudinal peer review function throughout, culminating in a public meeting to discuss the findings of the assessment and the assignment of ratings of clinical effectiveness and comparative value.

ICER has been purposely structured as a fully transparent organization that is able to engage with all key stakeholders in its appraisals while retaining complete independence in the formulation of its conclusions and the drafting of its reviews. ICER's academic mission is funded through a diverse combination of sources; funding is not accepted from manufacturers or private insurers to perform reviews of specific technologies. Since its inception, ICER has received funding from the following sources:

- The Agency for Healthcare Research & Quality (AHRQ)
- America's Health Insurance Plans (AHIP)
- Blue Shield of California Foundation
- Harvard Pilgrim Health Care
- HealthPartners
- Johnson & Johnson
- Merck & Co.
- The John W. Rowe Family Foundation
- Kaiser Health Plans
- The National Pharmaceutical Council
- United Health Foundation
- The Washington State Health Care Authority

More information on ICER's mission and policies can be found at:

www.icer-review.org

© 2009, Institute for Clinical and Economic Review

ACKNOWLEDGMENTS

We are grateful to Joseph Ladapo, MD, PhD, for allowing us to use published and unpublished models of CCTA use in the ED and outpatient settings as the structural basis for our economic evaluation, and for his technical assistance as we applied modifications to these models. We also wish to thank Katy Marttila for her ongoing technical assistance in preparing and disseminating this report.

EXECUTIVE SUMMARY

Introduction

Coronary computed tomographic angiography (CCTA) is a minimally invasive radiological technique used to provide images of the heart and surrounding vessels. CCTA has been suggested as an alternative or useful complementary approach to other non-invasive methods of diagnosing coronary artery disease (CAD). In particular, because of its ability to visualize coronary anatomy, CCTA has been suggested as a strategy to rule out significant CAD among patients at low or intermediate risk of significant disease, thereby giving greater reassurance than other non-invasive methods and potentially reducing the number of patients ultimately sent for invasive coronary angiography (ICA). However, uncertainty remains regarding several important issues:

1) The diagnostic accuracy of CCTA relative to ICA and other possible comparator diagnostic tests
2) The impact on patient outcomes and health care utilization of alternative diagnostic algorithms that integrate CCTA in different ways into the diagnostic pathways for patients with suspected CAD, both in the general outpatient setting and in the Emergency Department
3) The most appropriate target populations for CCTA, based on level of risk and symptoms
4) The potential negative impact of increased radiation exposure of CCTA
5) The impact of incidental findings that trigger further evaluation
6) The potential impact of CCTA on the thresholds for clinician testing for coronary artery disease among the general population
7) The budget impact and cost-effectiveness of integrating CCTA into diagnostic pathways for patients with suspected coronary artery disease

Given the possible benefits of introducing a widely available non-invasive option for CAD detection, the potential clinical and financial impact that broad adoption of CCTA would have on systems of care, and the uncertainty over the evidence on the net health benefits and appropriate use of CCTA, all health care decision makers will benefit from a formal appraisal of the comparative clinical effectiveness and comparative value of CCTA as a modality for diagnosis of coronary artery disease.

Coronary Artery Disease Diagnosis Alternatives

For many years the most precise and definitive method for the evaluation and diagnosis of coronary artery disease has been invasive coronary angiography (ICA). At the time of the procedure a catheter is inserted into an artery, usually the femoral blood vessel, and contrast dye is injected through the catheter. X-ray images are then captured and displayed on a video screen (a procedure known as fluoroscopy), and can be viewed either as images or in motion picture form. While complications from ICA are relatively infrequent, they can be significant, and include myocardial infarction, cardiac arrhythmia, stroke, hemorrhage, infection, trauma to the artery from hematoma or from the catheter, sudden hypotension, and reaction to the contrast medium (Gandelman, 2006). The procedure also

delivers a radiation dose in the range of 5-7 mSv, which is lower than most CCTA protocols but similar to that of CCTA when it is performed using dose-saving protocols or dual-source scanners.

In part because of the invasive nature of ICA and its concordant risks, alternative non-invasive tests also are utilized for evaluation of chest pain symptoms considered suggestive of CAD. The first of these technologies to gain widespread use was the stress electrocardiogram (EKG); the major alternatives are stress echocardiography and single-photon emission computed tomography (SPECT), also known as nuclear stress testing or myocardial perfusion imaging.

Stress echocardiograms (ECHO) produce images of the heart through the use of sound waves. The test allows for the evaluation of muscle function in different areas of the heart to identify weak or damaged areas of the muscle. This is done through a comparison of images at rest and under cardiac stress induced by exercise or pharmacologic means. Clinically, the test is simple to perform, relatively inexpensive, and easily accessible. However, the image quality is lower in obese patients and those with chronic lung disease, which can account for almost 30% of candidates (Miller, 2006). It is recommended for use in intermediate-to-high risk patients (Anthony, 2005).

SPECT imaging involves the use of a tracer radiopharmaceutical to highlight areas of decreased blood flow in the myocardium. Images are captured via a gamma camera, and may be reconstructed to create two or three-dimensional films. SPECT is often used in patients with intermediate-to-high risk for CAD. The accuracy of SPECT imaging has improved to the point that it is often used for prognostic use in addition to diagnosis. However, it has somewhat lower specificity in ruling out CAD in comparison to other diagnostic tests, and is not generally effective in detecting perfusion defects in patients with milder stenosis (Jeetley, 2006). SPECT also involves the use of contrast media and delivers a radiation dose somewhat higher in magnitude than that of ICA and CCTA (9-13 and 15-20 MSv for technetium and thallium isotopes respectively).

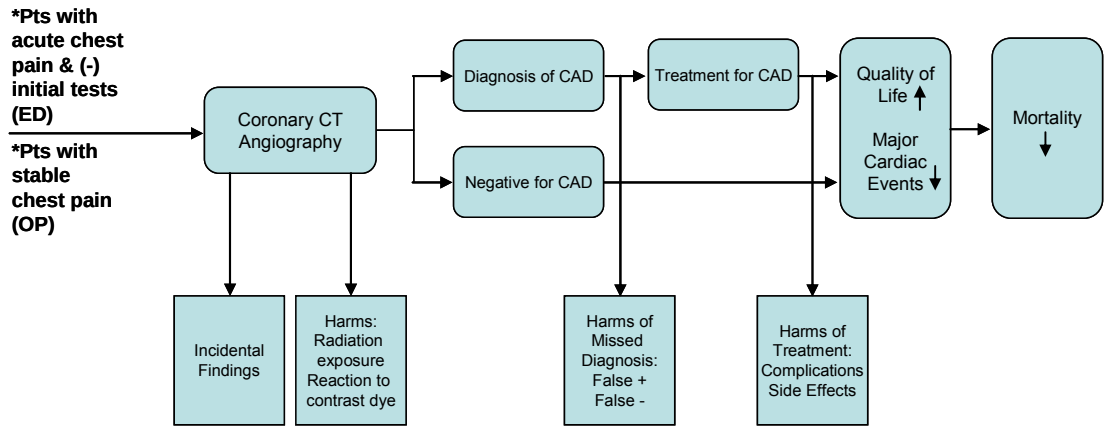
All of these alternative non-invasive diagnostic techniques measure in some way the functional impact on the heart of any underlying CAD. As noted above, none of the tests is perfect; each has the possibility of producing false positive and false negative results. Professional guidelines recognize all of these comparator techniques as appropriate initial investigations to evaluate possible CAD for most patients with stable symptoms (Gibbons, 2003).

Analytic Framework for Evaluation of CCTA

The analytic framework for this evaluation is shown in the Figure on the following page. As is the case for many diagnostic tests, there are no data directly demonstrating CCTA's beneficial impact on long-term morbidity and mortality, so judgments about the effectiveness of the intervention must rest almost exclusively upon consideration of the strength of sequential conceptual links. For this evaluation, the primary conceptual links

are those between detection of significant CAD, referral for appropriate treatment, major cardiovascular events, and mortality.

Analytic Framework: CCTA in ED and Outpatient Settings



Analytic Scope

CCTA provides different (visual) information than comparator non-invasive tests, and therefore simple comparisons of sensitivity and specificity against a gold standard (ICA) cannot provide adequate information on the downstream effects of CCTA on patient and clinician decision-making. There are both hypothetical benefits, such as reduced patient anxiety leading to reduced unnecessary follow-up testing, and hypothetical disadvantages, including the potential for overly aggressive management of mild-moderate levels of CAD. Because of the greater uncertainty in these potential effects of CCTA, the modeling effort of the ICER review provides analyses limited to the "*diagnostic phase*" (i.e., from patient presentation to diagnosis or rule-out of CAD) as well as traditional lifetime models.

CCTA Technical Evolution

CCTA is a technique in which a CT scanner is used to acquire multiple simultaneous tomographic sections ("slices") of the coronary arteries. At the time of this outpatient procedure, an IV is placed into a peripheral vein and a contrast dye is administered for the purposes of visually defining the arteries for the scan. Beta blockers may be given to the patient to slow the heart rate in order to prevent artifacts of heart motion that may affect image quality. The patient is positioned on the CT scanner and a large number of x-ray images are taken from multiple angles and reconstructed using computer software. Multi-detector row CT scanners contain rotating gantries that capture multiple images, or "slices". A 64-slice CCTA was introduced in 2004 and increased the number of captured images from the previous 16- and 32-slice technology. Improved spatial and temporal resolution from 64-slice machines has been found to shorten the time required to capture an image,

decreasing motion artifact as well as reducing the time to conduct the entire scan to approximately 8 seconds (Mowatt, 2008).

The 64-slice scanner has rapidly replaced earlier versions and is currently considered to be the community standard for CCTA. In 2007, 256- and 320-slice CT scanners became available, but it is unclear whether the greater resolution of these versions will provide clinically relevant advances to 64-slice machines. Dual source 64-slice scanners have also been introduced in which two scanners are mounted on the gantry at 90 degree angles (Matt, 2007). Dual source scanning is claimed by some to further decrease procedure time, reduce heart motion artifacts, and lower the effective radiation dose to the patient (Scheffel, 2006). In addition, as with any rapidly-evolving technology, it is unclear whether diagnostic performance as seen in studies conducted at highly-specialized academic centers will be representative of results obtained from use of CCTA in the general community.

This review included studies of the performance of CCTA in diagnosing CAD using scanners with 64-slice or higher resolution (including dual-source scanners). Guidance from the ICER Evidence Review Group suggested that 64-slice scanners were now widely available in the community and had become viewed as the standard for CCTA, and that literature on earlier-generation scanners would not be viewed as relevant by the clinical and patient communities.

Target Population for Consideration of Triage and Diagnosis of CAD

The accumulation of plaque that is characteristic of CAD typically gives rise to symptoms, such as chest pain and shortness of breath; in fact, the most important factors in determining CAD risk have been demonstrated to be age, gender, and the nature of chest pain (Diamond, 1979).

The relative effectiveness of any test used to detect CAD can be directly related to the perceived risk and/or underlying prevalence of significant disease. At the lowest levels of prevalence or risk, the benefits of accurate detection may be outweighed by the number of false positives generated by the test. Conversely, at the highest levels of prevalence or risk, patient populations are likely to benefit less from non-invasive diagnostic tests which will produce a relatively high rate of false negative results, and would instead benefit more from moving directly to definitive diagnostic testing and potential therapeutic intervention with ICA.

Following the guidance of the ICER Evidence Review Group (see section on Evidence Review Group starting on page 20) the target population for CCTA for this review was patients at *low-to-intermediate (10-30%) risk* of CAD, for the reasons given above. This review did not evaluate the performance of CCTA as a screening tool in very low-risk patients with non-specific chest pain or in asymptomatic patients. While the majority of diagnostic accuracy studies were conducted in relatively high-risk groups (i.e., patients already scheduled for ICA), we analyzed data separately by risk or pretest probability wherever feasible.

Evidence on Diagnostic Accuracy, Treatment Decisions, and Patient Outcomes
The available evidence on the impact of CCTA on clinician decision-making and patient outcomes is limited; nearly all available studies with these endpoints have been conducted in an ED setting; and, with the exception of one RCT, these studies have not prospectively compared the outcomes of "CCTA care" to the outcomes of standard care. The single published RCT compared a CCTA care strategy in the ED (n=99) to standard triage care alone (n=98) in an ED in Michigan (Goldstein, 2007); findings suggested that 67 (68%) patients in the CCTA care arm were identified with no CAD and were able to be rapidly discharged from the ED with no adverse outcomes over a 6-month follow-up period. More patients were sent to ICA in the CCTA care arm of the study (11 vs. 5), but 9 of 11 catheterizations proved "positive" in the CCTA care arm. CCTA was found to be time- and cost-saving due to a greater number of patients discharged immediately following a normal CCTA, a result that was echoed in another ED case series (Savino, 2006). In a second study of CCTA care in the ED, physicians in Israel evaluated 58 consecutive ED patients with standard triage care and made initial recommendations for disposition (Rubinshtein, 2007 [3]). Physicians were then given the patients' CCTA results, and the impact on final disposition decisions and patient outcomes suggested that CCTA findings prevented unnecessary hospitalization or invasive treatment in 40-45% of patients.

There are two important considerations in these ED studies. First, they are small studies, and in both the overall risks of acute coronary syndrome and cardiac events were very low. As one of the authors notes, the lack of negative outcomes among CCTA-negative patients cannot be taken as conclusive evidence of the true incidence of false positive and false negative CCTA findings. These studies also highlight how critical the underlying prevalence and distribution of CAD is in understanding the relative effectiveness of CCTA as a diagnostic and triage modality.

In the outpatient setting, where the interest in the use of CCTA has been focused on the evaluation of patients with stable chest pain symptoms who are at low-to-intermediate risk of significant CAD, the few published studies to date that have directly and prospectively measured the impact of CCTA on clinical decision-making or on patient outcomes have not included any controlled comparison arm of patients managed without CCTA. The majority of available literature on 64-slice CCTA is limited to small, single-center studies of diagnostic accuracy compared to ICA, typically among consecutive patients at relatively high risk of CAD who are already scheduled to undergo ICA. This body of evidence has expanded rapidly from 2005-2008, and the findings are relatively consistent. Our pooled estimate (from meta-analysis of 34 studies) of the sensitivity of CCTA for significant CAD is high: 97%; 95% CI, 96%, 98%. This sensitivity compares favorably to estimates for alternative non-invasive techniques including stress ECHO (76-94%) and SPECT (88-98%) (Garber, 1999).

The specificity of CCTA can be calculated in two ways based on how scans with "non-diagnostic" segments are treated. When patients with non-diagnostic CCTA results were counted as false-positives, pooled specificity from the ICER meta-analysis was 82% (95% CI: 79%, 84%); when such patients were excluded from analyses (as they were in most of the

studies we analyzed), specificity was calculated to be 87% (95% CI: 85%, 89%). This range for specificity is also comparable or superior to estimates for other non-invasive techniques: 88% for stress ECHO and 77% for SPECT (Garber, 1999). A significant degree of heterogeneity was found in the specificity estimates; in exploratory analyses, the only significant source of heterogeneity was found to be age, with studies of older patients producing more variable findings. However, because pooled estimates from studies of younger populations were essentially identical to the overall meta-analytic findings, no further adjustment to the overall estimates was required.

Regardless of the level of confidence in diagnostic accuracy findings, sensitivity and specificity estimates by themselves cannot suggest how CCTA results would affect clinical decision-making or patient outcomes. For one thing, CCTA results in practice are not interpreted in a binary fashion. Many patients will have "moderate" stenosis (20%-70%) in one or more arteries. One of the important unanswered questions about CCTA is the clinical significance and the impact on clinical decision-making of visual identification of moderate stenosis. Prior to CCTA these patients would have undergone either non-invasive tests, which would have evaluated *functional* signs of CAD without any visual image, or these patients would have been sent directly for ICA. How CCTA would affect the diagnoses and pattern of care for patients with "moderate" stenosis is a controversial topic. Some authors have postulated use of CCTA would increase testing rates based on an "oculostenotic reflex," the compulsion that cardiologists might feel to aggressively treat any occlusion they see (Lin, 2007; Topol, 1995). Others have hypothesized that visualization of moderate stenosis, particularly at the lower end of the 20%-70% range, will prove reassuring to clinicians and patients, reducing repeat testing and inappropriately aggressive therapy (Valenza, 2006). Unfortunately, there are no published data with which to evaluate how clinical decision making for patients with moderate stenosis in the outpatient setting changes with the integration of CCTA into practice.

There are several other important issues to note regarding the evidence on diagnostic accuracy. The prevalence of underlying CAD is quite high in many of the accuracy studies (mean of 59% in the studies analyzed), raising questions about the applicability of study results from these populations to those including a preponderance of "low-to-intermediate" risk. Although published data suggest that CCTA's accuracy is unaffected by the extent and distribution of CAD in the population, the absolute number of indeterminate and false positive results from CCTA would be higher in any population with a lower true prevalence of disease.

And finally, given the long-term progression inherent in CAD, and the uncertainties surrounding its natural history, the lack of published evidence makes it difficult to judge the magnitude of the benefits of reductions in false negative and false positive diagnoses. There is no published evidence to judge the outcomes of patients with initially false negative stress ECHO, SPECT, or CCTA results. Some will suffer a preventable cardiac event; others will return in the near future for further evaluation, be correctly diagnosed, and will be treated appropriately with little negative impact on health outcomes. Similarly, the balance of net harms and benefits is unknown for patients receiving a false positive

diagnosis of CAD with CCTA or any of the non-invasive testing strategies. These patients will receive the "harms" of unnecessary medical therapy in the short term, but the balance of these harms against the potential benefits in patients who would develop CAD over time is unknown.

Harms

Review of the evidence confirmed that CCTA is a safe procedure, with the only immediate complication being reactions to contrast media; the reported rates of serious contrast reactions or induced nephropathy has been very low for the technologies that require contrast, and the rate of reactions requiring serious intervention (e.g., dialysis, hospitalization) has been even lower.

To place the effective radiation dose received from CCTA in some context, the average reported range of radiation in our sampled studies is listed in the table below along with typical doses from other tests and exposures to x-rays. Note that the doses received from ICA are similar to those at the lower end of the reported range for CCTA, while the range of SPECT doses are similar to those at the higher end of the reported range for CCTA:

Radiation exposure scenario	Approximate effective dose (mSv)
Chest x ray	0.02
Round-trip flight, New York-Seattle	0.06
Low-dose CT colonography	0.5-2.5
Lumbar spine x-ray	1.3
Head CT	2.0
Single-screening mammogram (breast dose)	3.0
Annual background dose caused by natural radiation	3.0/yr
CCTA (lower reported range)	**2.0-8.0**
Invasive coronary angiography	***5.0-7.0***
Adult abdominal CT scan	10.0
Single photon emission CT (SPECT): Technetium	***9.0-13.0***
CCTA (higher reported range)	**12.0-14.0**
Typical dose to A-bomb survivor at 2.3 km distance from ground zero Hiroshima	13.0
SPECT: Thallium	***15.0-20.0***
Annual radiation worker annual exposure limit	20.0/yr
Annual exposure on international space station	170.0/yr

Sources: Brenner, 2005; FDA [www.fda.gov/cdrh/ct/risks.html]; ICER CCTA systematic review; Van Gelder 2004, Mettler 2008, Shuman 2008; Earls 2008; Husmann 2008 [2].

The potential for harm from radiation is more difficult to assess given the uncertainty around the relationship between low-level radiation exposure and cancer risk as well as whether an exposure threshold exists above which excess risk is realized. One published empirical attempt to quantify the lifetime attributable risk for cancer estimated that it is

0.22% and 0.08% in women and men aged 60 years respectively; prospective EKG gating would be expected to reduce this risk by about 35% (Einstein, 2007). Aggressive attempts are being made to reduce radiation dose during CCTA, with varying degrees of success; still, consideration of CCTA's radiation dose is important, particularly in light of the possible exposure from other tests along the diagnostic pathway (e.g., SPECT, ICA).

Incidental Findings

The relative benefits and harms of incidental findings on CCTA are also difficult to judge empirically. Studies suggest that approximately 40-80% of patients will have an extra-coronary finding of some kind on CCTA, and 5-20% of patients would have a finding deemed clinically important enough for further evaluation. Were CCTA to be adopted broadly, this rate of extra-coronary findings would generate significant numbers of patients requiring further investigation. When investigated, some of these findings will be judged to have brought clinical benefit to the patient, most often by detection of a pulmonary malignancy or embolism, or possibly diagnosis of an abdominal or thoracic aortic aneurysm. However, findings from the few studies that have examined this question suggest that the proportion of patients receiving some clinical benefit is very low, while additional risks, anxieties, and costs are generated by follow-up investigations (Onuma, 2006; Cademartiri, 2007 [4]). The results of our analyses suggest that the additional costs of following patients for pulmonary nodules alone are approximately $100 per patient undergoing CCTA. From both a clinical and a health systems perspective this is one of the most important uncertainties regarding CCTA. The determination of net health benefit for CCTA may hinge on decision-makers' interpretation of the boundaries of risk, benefit, and cost of extra-coronary findings. As highlighted previously, this is but one of the key uncertainties around CCTA's diffusion in clinical practice; for example, if CCTA's use expands to low-risk populations in which the balance of true and false positives is less certain, the uncertainties around incidental findings take on added significance.

Clinical Effectiveness Results from ICER Decision Analytic Models

Because the clinical scenarios and patient populations related to CCTA use differ substantially between the ED and the outpatient settings, we decided to build two separate models that could help evaluate the likely impact of CCTA compared to alternative diagnostic strategies in these two settings. Due to lack of reliable data and no consensus among clinical and policy experts, neither model explicitly includes the potential benefits, harms, or costs of incidental findings or radiation exposure; however, in a post hoc analysis, an attempt is made to quantify the cost impact from short-term follow-up of incidental findings in the ED.

Triage of Patients in the ED

The model evaluating CCTA for patients with acute chest pain in the ED setting follows the algorithm of the RCT by Goldstein (Goldstein, 2007) but with one important difference. As with the Goldstein protocol, patients are at low-to-intermediate risk of an acute coronary syndrome, with negative initial serum enzyme tests and no significant EKG elevations. But Goldstein's trial only randomized patients who had completed a second negative serum enzyme test at 4 hours. Our model assumes that patients in the CCTA arm do not wait for a

second serum test before being sent for CCTA. In the CCTA pathway all patients receive CCTA immediately, with subsequent triage determined by CCTA results. Standard of care (SOC) in our model includes admission to an ED observation unit to await final serum enzyme tests, followed by SPECT if final enzymes are also negative; in an alternative scenario, we replace SPECT with stress ECHO as the standard stress-test modality. Details of the model are available in Section 8.

Table ES1 below depicts the ED model results for a cohort of 1,000 55-year old men. The left hand column shows the result if all patients had undergone the SOC strategy and the right hand column depicts the results if the identical 1,000 patients had all undergone the CCTA strategy. Among the notable differences between CCTA and SOC are the number of patients sent immediately home without requirement for extended ED observation (567 vs. 0, data not shown); the number of false negatives (16 vs. 63), the number of false negatives that represented "missed" cases of acute coronary syndrome (5 vs. 18), the number of patients ultimately referred for ICA (327 vs. 434), and the number of patients sent for ICA who are found to have normal coronary arteries on ICA (74 vs. 228).

The results of our model are consistent with other published cost-effectiveness analyses in suggesting that when used as part of a triage strategy for low-to-intermediate risk chest pain patients in the ED, CCTA will allow more rapid discharge of nearly half of all patients and decrease the number of false negative diagnoses while reducing the number of angiographies compared to the current standard of care. However, these findings contrast with the results from Goldstein's RCT, which found a higher rate of ICA in the CCTA arm. We believe this seeming contradiction is primarily driven by two modeling assumptions: 1) a higher prevalence of CAD in the patient cohort; and 2) both arms begin with patients *prior* to a second negative serum enzyme test, increasing the number who "rule-in" for acute coronary syndrome. In addition, the number of patients in the Goldstein study is relatively small, and it is difficult to determine whether the higher CCTA rate found was a true consequence of the care pathway or due to chance.

Table ES1: Base case results of ED model

Outcomes (per 1,000)	SOC	CCTA
True positive	206	253
True negative	731	731
False negative	63	16
False negative w/ACS	18	5
Referred for ICA	434	327
ICA negative results	228	74
ICA related deaths	0.04	0.03
Incidental findings	0	138

Notes: SOC: standard of care; ACS: acute coronary syndrome

Evaluation of Stable Chest Pain in the Outpatient Setting
The model evaluating CCTA as a tool for evaluating stable chest pain in the outpatient setting follows the CAD treatment recommendation derived from the recent COURAGE trial (Boden, 2007) and thus requires that the diagnostic tests not only identify stenoses correctly but also differentiate between 3-vessel/left main artery disease and 1- or 2-vessel disease.

The base case population consisted of 55 year-old men with stable chest pain and with either low (10%) or intermediate (30%) prevalence of underlying significant CAD -- one or more vessels with occlusion ≥70% or left main occlusion at ≥50%. We considered 8 different strategies, alone and in combination, in order to capture a wide range of management approaches for evaluating patients with stable chest pain and a low-to-intermediate risk of CAD:

1. Coronary Computed Tomographic Angiography (CCTA)
2. Stress-Echocardiography (Stress-ECHO)
3. Stress- Single Photon Emission Computed Tomography (Stress-SPECT)
4. CCTA followed by Stress-ECHO
5. Stress-ECHO followed by CCTA
6. CCTA followed by Stress-SPECT
7. Stress-SPECT followed by CCTA
8. Stress-ECHO followed by Stress-SPECT

Table ES2 on the following page depicts the base case model results for 1,000 55-year old men with an underlying CAD prevalence of 30%. Each column represents the results if all patients had undergone the specific screening strategy.

The model results indicate that there are important trade-offs to consider when comparing these strategies. There is no single, simple axis of "effectiveness." For example, "*CCTA alone*" has the highest number of true positives at 288 and the lowest number of false negatives at 8 (2 of whom have 3-vessel or left main disease) among all strategies, followed by "*SPECT alone*" which has 271 true positives and 25 false negatives. But CCTA strategies introduce the issue of incidental findings, estimated to require follow-up among 13.8% of all patients screened. CCTA (and SPECT) strategies also carry radiation exposure risks for all patients. By scanning and comparing the columns in the Table decision-makers can weigh the value they ascribe to these different aspects of the outcomes associated with various diagnostic strategies. A Table showing results for a lower-risk population with a 10% prevalence of CAD, shown in Section 8 of the review, also demonstrates how these various outcomes shift importantly with the underlying prevalence of disease in the population.

Table ES2: Diagnostic results in the Outpatient Setting (30% CAD prevalence)

Estimates	CCTA	SPECT	SECHO	CCTA -> SPECT	SPECT -> CCTA	CCTA -> SECHO	SECHO -> CCTA	SECHO -> SPECT
True positive	288	271	245	266	265	245	239	228
False positive	86	149	74	23	26	11	19	33
True negative	618	556	631	682	679	694	686	672
False negative	8	25	50	29	31	51	56	68
False negative w/3-v or LM disease	2	1	4	2	1	2	4	4
Referred for ICA	107	160	195	106	90	118	85	105
ICA-negative results	21	61	89	7	5	11	4	12
ICA related deaths	0.11	0.17	0.20	0.11	0.09	0.12	0.09	0.11
Exposed to radiation	1000	1000	195	1000	1000	1000	408	408
Incidental findings requiring f/u	138	0	0	138	57	138	47	47
Total costs/patient [excluding all f/u costs, $]	760	1,204	837	1,002	1,203	886	694	850

Notes: CCTA: coronary computed tomographic angiography; SPECT: single photon emission computed tomography; SECHO: stress echocardiogram; 3-v: 3-vessel coronary artery disease; LM: coronary artery disease of the left main artery; ICA: invasive coronary angiography; f/u: follow-up

Summary of Findings of Comparative Value

ED Setting

We performed cost-effectiveness analyses using the decision analytic models described above. According to the base case results of the ED model, CCTA is cost-saving, with about $719 in savings per patient in comparison to SOC. Taking into account the additional follow-up costs for the 14% of patients who undergo CCTA and have incidental findings, the cost-savings are reduced to about $619, but remain in favor of CCTA. The following numbers represent the base case analysis and compare CCTA in addition to standard triage care to standard care alone:

- Cost of CCTA= $466

- CCTA cost savings relative to standard care (includes CCTA, ED triage, observation, cath lab) = $719

- CCTA cost savings w/incidental findings f/u costs = $619

- Threshold CCTA cost for cost savings in the ED = $1,185

When the diagnostic modality in the SOC pathway was changed to stress ECHO, the number of true positives decreased, as SPECT is a more sensitive test than stress ECHO. However, stress ECHO has higher specificity, which resulted in a decrease in the numbers of patients referred for ICA and ICA-negative results. Based on these tradeoffs, as well as the increased test costs with SPECT ($765 vs. $300 for stress ECHO), a CCTA-based strategy remains cost saving, with estimated savings of $314 per patient vs. patients triaged using stress ECHO.

Outpatient Evaluation: Diagnostic Phase

The outpatient model was used to evaluate testing costs of the diagnostic phase, extending up through and including possible ICA but not beyond. Table ES2 on the previous page includes, in the final row, the average diagnostic costs per patient generated by the base case model at 30% CAD prevalence. The CCTA alone strategy was found to be less expensive ($760 per patient) than all other diagnostic strategies except for Stress ECHO followed by CCTA ($694 per patient). It should be noted again that these cost estimates do not include the subsequent costs of evaluation for incidental findings, which we estimate averages $100 per patient sent for CCTA.

Outpatient Evaluation: Lifetime Model

A formal cost-effectiveness analysis comparing all the outpatient evaluation strategies was performed considering a lifetime horizon for cardiac outcomes and costs. Strategies were similar in effectiveness, as about 2 weeks of quality-adjusted life expectancy separated the most and least effective strategies. As compared to stress ECHO, CCTA alone was more expensive but also more effective, and therefore an incremental cost-effectiveness ratio for CCTA alone was calculated:

- Cost per QALY* gained vs. Stress ECHO = $13,100

*QALY = Quality adjusted life year

CCTA alone was more effective and less costly than SPECT alone. In addition, all of the combination strategies evaluated were less effective than single-test strategies. Finally, at a cost of $248 or less, CCTA would be a dominant (i.e., cost-saving) strategy relative to stress ECHO.

Note that, when a 10% CAD prevalence is considered, the relative costs of strategies involving CCTA increase due to the greater number of false-positive results generated and lessening of differences in the absolute number of false negatives between strategies.

CCTA's profile as compared to stress ECHO remains essentially unchanged (cost/QALY of $17,000); however, while still more costly, SPECT alone is more effective than CCTA, at a cost/QALY of $82,300 relative to CCTA. In addition, the combination of SPECT followed by CCTA appears more effective and less costly than CCTA alone at this level of disease prevalence.

ICER Evidence Review Group Deliberation

The ICER Evidence Review Group deliberation (see section starting on page XX for membership and details) focused on many important issues regarding the evidence provided by the ICER review. Major points of discussion are shown in the numbered points below.

1) *Following ICER's conduct of meta-analyses of diagnostic accuracy based on single-center studies, results of two major multi-center studies (ACCURACY and CORE 64) became available in the literature. Findings from these studies differed substantially – the ACCURACY results were similar to ICER's findings, while the CORE 64 results showed lower sensitivity and higher specificity.*
   The ERG discussed these results in detail; one hypothesis for the difference in findings was that CORE 64 was an international study, and there might have been more variability in CCTA practices and diagnostic thresholds. One ERG member mentioned potential inconsistencies at one of the dominant CORE 64 sites, although this was not described in the publication. In any event, there was consensus that these two studies should be included in the meta-analysis and possibly weighted in some way over single-center studies. The inclusion of these studies did not materially change the original meta-analysis results, as now discussed in the report; details of the studies themselves have been added to the report as well.

2) *Because the evidence of diagnostic accuracy is driven by small, single-center studies, exploratory analyses should be conducted to ascertain publication bias.*
   Examinations of both heterogeneity and publication bias have now been undertaken and added to the body of the review. For the former, threshold analyses and meta-regression were undertaken to understand the sources of heterogeneity; for the latter, efforts were made to eliminate duplicative results and identify significant unpublished research.

3) *The discussion of the results should include the concept of "spectrum bias"; i.e., the possibility that examination of CCTA accuracy in populations with high CAD prevalence and/or severe disease might over-estimate sensitivity and specificity.*
   This has been added to the discussion of the systematic review findings, as have the results of analyses previously run to address this issue: (a) comparison of test characteristics between studies that included patients with known CAD vs. those that did not; and (b) summarization of studies that stratified findings by CAD risk or pretest probability.

4) *Because CCTA is not indicated in certain circumstances (e.g., high levels of coronary calcium), some attempt to quantify the proportion of candidates for non-invasive CAD testing in each setting for whom CCTA would be appropriate.*
   These statistics have been added to the description of CCTA technology.

5) *In discussions of the potential harm from radiation dose for CCTA and other radiation-based technologies, some mention should be made of the notion that reported rates are "moving targets", and that active efforts are underway to reduce radiation dose from all of these technologies. In addition, age at time of exposure is an important consideration for all of these technologies.*
   The report and discussion of harms has been revised to reflect these constructs.

6) *While incidental findings remain a controversial topic with CCTA, a joint registry involving several medical and imaging societies is planned in part to address long-term follow-up and outcomes from extra-coronary findings on CCTA.*

7) *Changes were recommended for the economic model of CCTA in the ED setting to better reflect clinical practice: (a) instead of immediately discharging 50% of patients with mild/moderate stenosis on CCTA and sending 50% into standard-care triage, the percentages should be adjusted to be 80% and 20% respectively; and (b) in the standard-care arm, 20% of patients with a second negative troponin test should be immediately discharged, and the remaining 80% should receive a stress test.*
   These changes have been made; this structure is now considered the new "basecase" for the ED model.

8) *While the diagnostic phase results are of interest, more data should be made available; specifically, for the ED model, the proportion of false negatives that were missed cases of acute coronary syndrome, and for the outpatient model, the proportion of the same with 3-vessel or left main disease should be disclosed.*
   We have modified the diagnostic phase results to reflect these data.

9) *Some disaggregation of the cost findings, particularly with respect to lifetime results for the outpatient model, would be valuable to understand the major drivers of the findings.*
   The report has been expanded to include discussion of this issue.

10) *The assumption of independent test performance in the model is a limitation, in that there is likely some degree of complementarity in multi-test strategies for CAD.*
   As discussed during the meeting, the project timeframe did not allow for complex modeling the complementary nature of multi-test strategies, although there is some evidence that CCTA's visual aspects do complement the functional results from other tests. This has been noted in a new limitations section in the report.

<u>Discussion of ICER Integrated Evidence Ratings</u>

The specific discussion of the assignment of ICER ratings for comparative clinical effectiveness and for comparative value were conducted separately for the ED and outpatient settings respectively. In the ED setting, the majority (8/11) of participants felt

that the evidence was sufficient to rate CCTA as at least "Comparable" to standard triage care. Some ERG members felt that the evidence base, while promising, was still too thin to label CCTA at a level higher than "Unproven with Potential", while others felt that the potential for avoiding unnecessary angiography and efficient ED triage was enough to label CCTA's net health benefits "Incremental". Most of the ERG participants (8/11) also agreed that the cost savings with CCTA in the ED model translated to a comparative value rating of "High"; the remainder of participants rated the technology as "Reasonable/Comparable" or on the continuum between these two levels.

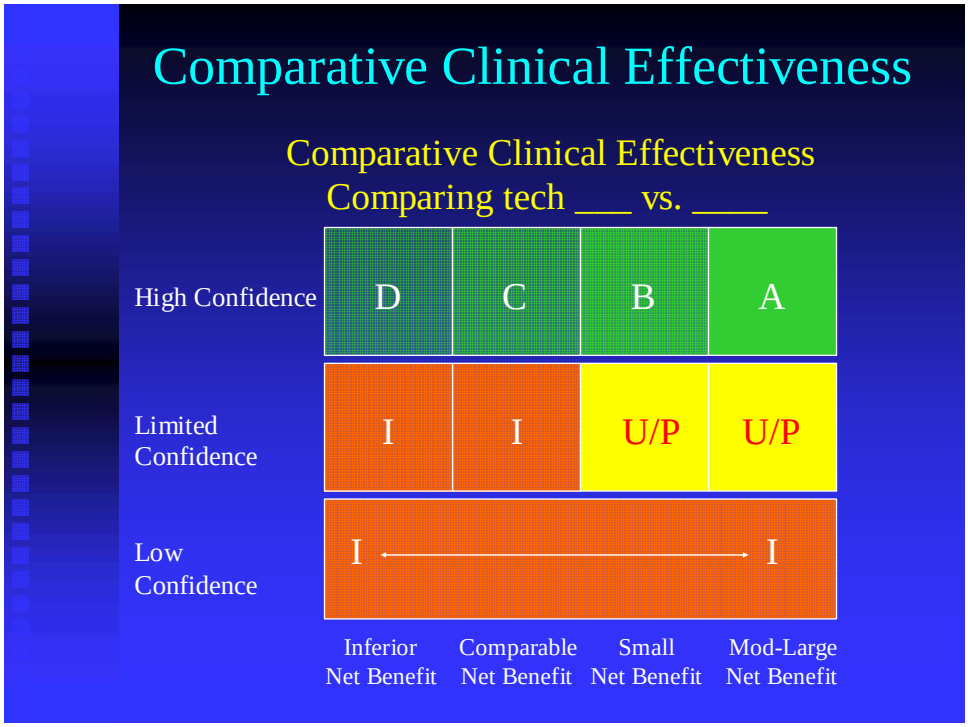
There was recognition that the evidence base for patient outcomes of CCTA in the outpatient setting was not as solid, and this was reflected in the ratings of comparative clinical effectiveness. While 4 of 11 ERG members felt that CCTA should be rated as at least "Comparable" to other non-invasive strategies, an equal number felt that the technology was still "Unproven" or the evidence was "Insufficient". Two additional participants felt that the rating was somewhere between "C" and "U/P", and one felt that CCTA's superior test characteristics provided "Incremental" benefit. Regarding comparative value, the group was unanimous in presenting CCTA's value as "Reasonable/Comparable" to other non-invasive strategies.

The input of the ERG is advisory to ICER; the ultimate rating is made after independent discussion and reflection on the entirety of the review as well as associated meetings. Background on the ICER rating methodology is shown on the following pages, with the final ICER ratings immediately afterward.

Methodology: ICER Integrated Evidence Rating™

Comparative Clinical Effectiveness

The ICER Integrated Evidence Rating™ combines a rating for comparative clinical effectiveness and a rating for comparative value. The clinical effectiveness rating arises from a joint judgment of the level of confidence provided by the body of evidence and the magnitude of the net health benefit -- the overall balance between benefits and harms. This method for rating the clinical effectiveness is modeled on the "Evidence- Based Medicine (EBM) matrix" developed by a multi-stakeholder group convened by America's Health Insurance Plans. This matrix is depicted below:



A = "Superior" [High confidence of a moderate-large net health benefit]
B = "Incremental" [High confidence of a small net health benefit]
C = "Comparable" [High confidence of a comparable net health benefit]
D = "Inferior" [High confidence of an inferior net health benefit]
U/P = "Unproven with Potential " [Limited confidence of a small or moderate-large net health benefit

This category is meant to reflect technologies whose evidence provides:

1) High confidence of *at least* comparable net health benefit
2) Limited confidence suggesting a small or moderate-large net health benefit

I = "Insufficient" The evidence does not provide high confidence that the net health benefit of the technology is at least comparable to that provided by the comparator(s).

Confidence
The vertical axis of the matrix is labeled as a degree of confidence with which the magnitude of a technology's comparative net health benefit can be determined. This operational definition of confidence thus is linked to but is not synonymous with the overall validity, consistency, and directness of the body of evidence available for the assessment. ICER establishes its rating of level of confidence after deliberation by the Evidence Review Group, and throughout ICER follows closely the considerations of evidentiary strength suggested by the Effective Health Care program of the Agency for Health Research and Quality (AHRQ) (www.effectivehealthcare.org) and the GRADE working group (www.gradeworkinggroup.org).

High Confidence:
An assessment of the evidence provides high confidence in the relative magnitude of the net health benefit of the technology compared to its comparator(s).

Limited Confidence:
There is limited confidence in the assessment the net health benefit of the technology. Limited confidence implies that the evidence is limited in one or more ways so that it is difficult to estimate the net health benefit with precision. ICER's approach considers two qualitatively different types of limited confidence. First, there may be limited confidence in the magnitude of any net health benefit, but there is high confidence that the technology is *at least* as effective as its comparator(s). The second kind of limited confidence applies to those technologies whose evidence may suggest comparable or inferior net health benefit and for which there is not nigh confidence that the technology is at least comparable. These two different situations related to "limited confidence" are reflected in the matrix by the different labels of "Unproven with Potential" and "Insufficient."

Limitations to evidence should be explicitly categorized and discussed. Often the quality and consistency varies between the evidence available on benefits and that on harms. Among the most important types of limitations to evidence we follow the GRADE and AHRQ approaches in highlighting:

1. Type of limitation(s) to confidence
    a. Internal validity
        i. Study design
        ii. Study quality
    b. Generalizability of patients (directness of patients)
    c. Generalizability of intervention (directness of intervention)
    d. Indirect comparisons across trials (directness of comparison)
    e. Surrogate outcomes only (directness of outcomes)
    f. Lack of longer-term outcomes (directness of outcomes)
    g. Conflicting results within body of evidence (consistency)

Low Confidence:
There is low confidence in the assessment of net health benefit and the evidence is insufficient to determine whether the technology provides an inferior, comparable, or better net health benefit.

Net Health Benefit
The horizontal axis of the comparative clinical effectiveness matrix is "net health benefit." This term is defined as the balance between benefits and harms, and can either be judged on the basis of an empiric weighing of harms and benefits through a common metric (e.g. Quality Adjusted Life-Years, or "QALYs"), or through more qualitative, implicit weightings of harms and benefits identified in the ICER appraisal. Either approach should seek to make the weightings as explicit as possible in order to enhance the transparency of the ultimate judgment of the magnitude of net health benefit.

Whether judged quantitatively or qualitatively, there are two general situations that decision-making groups face in judging the balance of benefits and harms between two alternative interventions. The first situation arises when both interventions have the same types of benefits and harms. For example, two blood pressure medications may both act to control high blood pressure and may have the same profile of side effects such as dizziness, impotence, or edema. In such cases a comparison of benefits and harms is relatively straightforward. However, a second situation in comparative effectiveness is much more common: two interventions present a set of trade-offs between overlapping but different benefits and harms. An example of this second situation is the comparison of net health benefit between medical treatment and angioplasty for chronic stable angina. Possible benefits on which these interventions may vary include improved mortality, improved functional capacity, and less chest pain; in addition, both short and long-term potential harms differ between these interventions. It is possible that one intervention may be superior in certain benefits (e.g. survival) while also presenting greater risks for particular harms (e.g. drug side effects). Thus the judgment of "net" health benefit of one intervention vs. another often requires the qualitative or quantitative comparison of different types of health outcomes.

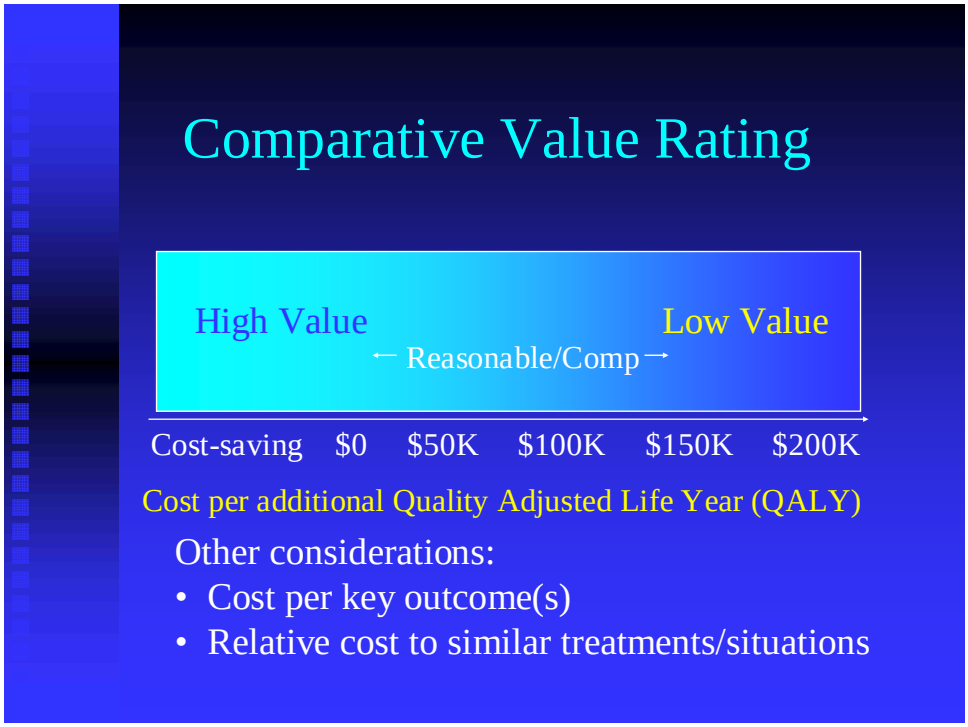
Since net health benefit may be sensitive to individual patient clinical characteristics or preferences there is a natural tension between the clinical decision-making for an individual and an assessment of the evidence for comparative clinical effectiveness at a population level. ICER approaches this problem by seeking, through the guidance of its scoping committee, to identify a priori key patient subpopulations who may have distinctly different net health benefits with alternative interventions. In addition, the ICER appraisal will also seek to use decision analytic modeling to identify patient groups of particular clinical characteristics and/or utilities which would lead them to have a distinctly different rating of comparative clinical effectiveness.

The exact boundary between small and moderate-large net benefit is subjective and ICER does not have a quantitative threshold. The rating judgment between these two categories is guided by the deliberation of the Evidence Review Group.

© 2009, Institute for Clinical and Economic Review

Comparative Value

There are three categories of value: high, reasonable or comparable, and low. The ICER rating for comparative value arises from a judgment that is based on multiple considerations. Among the most important is the incremental cost-effectiveness of the technology being appraised The most commonly used metric for an assessment of cost-effectiveness is the quality adjusted life year, or QALY. This measure adjusts any improvement in survival provided by a technology by its corresponding impact on the quality of life as measured by the "utilities" of patients or the public for various health states. While ICER does not operate within formal thresholds for considering the level at which a cost per QALY should be considered "cost-effective," the assignment of a rating for comparative value does build upon general conceptions of ranges in which the incremental cost-effectiveness ratio can be generally assumed to indicate relatively high, reasonable, and low value compared to a wide range of health care services provided in the US healthcare system. These broad ranges and shown in the figure below. Details on the methodology underpinning the design and presentation of cost-effectiveness analyses within ICER appraisals is available on the ICER website at www.icer-review.org.



Although the cost per QALY is the most common way to judge the cost-effectiveness of alternative medical interventions, ICER also considers the sub-component parts of the QALY, including the cost per key clinical benefits. Additional data and perspectives are also considered whenever possible, including potential budget impact, impact on systems of care and health care personnel, and comparable costs/CEA for interventions for similar clinical conditions.

Integrated Ratings

The ICER Integrated Evidence Rating™ combines the individual ratings given for comparative clinical effectiveness and comparative value. The overall purpose of the integrated ratings is to highlight the separate considerations that go into each element but to combine them for the purposes of conveying that clinical benefits provided by technologies come at varying relative values based on their cost and their impact on the outcomes of care and the health care system.

ICER Integrated Evidence Rating™: CCTA vs. Standard ED Triage Care

The Comparative Clinical Effectiveness of CCTA for triage of patients with acute chest pain and at low to intermediate risk of acute coronary syndromes in an ED setting is rated as:

- **C --- Comparable**

The Comparative Value of CCTA for triage of patients with acute chest pain in an ED setting is rated as:

- **a --- High***

The Integrated Evidence Rating = Ca*

*** Within assumptions of the economic analysis, including reimbursed price of CCTA assumed to = $466**

ICER Integrated Evidence Rating™: CCTA vs. Standard ED Triage Care

Comparative Clinical Effectiveness	a High	b Reasonable/Comp	c Low
Superior: A	Aa	Ab	Ac
Incremental: B	Ba	Bb	Bc
Comparable: C	CCTA=Ca	Cb	Cc
Unproven/Potential: U/P	Ua	Ub	Uc
Insufficient: I	I	I	I

Comparative Value

ICER Integrated Evidence Rating™: CCTA vs. Alternative Outpatient Strategies for Stable Chest Pain

The Comparative Clinical Effectiveness of CCTA for assessment of outpatients without signs or symptoms of unstable chest pain and at low to intermediate risk of significant coronary artery disease is rated as:

- **U/P – Unproven but with Evidence of Potential Net Benefit**

The Comparative Value of CCTA for assessment of outpatients presenting with stable chest pain is rated as:

- **b --- Reasonable/Comparable***

The Integrated Evidence Rating = Ub*

*** Within assumptions of the economic analysis, including reimbursed price of CCTA assumed to = $466**

ICER Integrated Evidence Rating™: CCTA vs. Alternative Strategies for Stable Chest Pain

Comparative Clinical Effectiveness	a High	b Reasonable/Comp	c Low
Superior: A	Aa	Ab	Ac
Incremental: B	Ba	Bb	Bc
Comparable: C	Ca	Cb	Cc
Unproven/Potential: U/P	Ua	CCTA=Ub	Uc
Insufficient: I	I	I	I

Comparative Value

Evidence Review Group Members

The Evidence Review Group (ERG) is an independent group brought together by ICER and composed of academic experts, patients, clinicians, epidemiologists, ethicists, and medical policy representatives of stakeholder groups including health plans and manufacturers.

The purpose of the ERG is to guide and help interpret the entire appraisal process. Members of the ERG are first convened to function as a "scoping committee" for the appraisal. During this phase the key questions for the appraisal are outlined, including elements such as the appropriate comparator technologies, patient outcomes of interest, patient subpopulations for which clinical and cost-effectiveness may vary systematically, time horizon for outcomes, and key aspects of the existing data that must be taken into account during the appraisal. The ERG may be divided into sub-committees that advise the ICER appraisal team at the mid-point of the appraisal on the early findings and challenges encountered.

At the final ERG meeting, members are asked to declare any interests in the technology or its comparator(s). The ERG meeting allows for in-depth deliberation on the findings of the ICER appraisal document and provides an opportunity for comment on the determination of the ICER integrated evidence rating. Although the ERG helps guide the final determination of the ICER Integrated Evidence Rating™, the final rating is ultimately a judgment made by ICER, and individual members of the ERG should not be viewed in any way as having endorsed this appraisal.

ERG Participant Name	Potential Influences on Expertise
Robin Cisneros Director, Medical Technology Assessment and Products The Permanente Foundation (Kaiser)	Reviews evidence on medical technology for payer
G. Scott Gazelle, MD, MPH, PhD Director, Institute for Technology Assessment Professor of Radiology Professor of Health Policy & Management Massachusetts General Hospital & Harvard Medical School	None
Alan Go, MD Assistant Director, Clinical Research Senior Physician, Division of Research Kaiser Permanente, Northern California	Not present at meeting

Mark Hlatky, MD Professor of Health Research & Policy Professor of Medicine Stanford University	Consulting relationships with GE Healthcare and Blue Cross Blue Shield Association
Udo Hoffmann, MD, MPH Director, Cardiac MR PET CT Program Associate Professor of Radiology Massachusetts General Hospital & Harvard Medical School	None
Leah Hole–Curry, JD Director, Health Technology Assessment State of Washington Health Care Authority	Not present at meeting
Robert Honigberg, MD, MBA Chief Medical Officer Global Technology Medical Organization, GE Healthcare	Employed by GE Healthcare
Jill Jacobs, MD Chief, Cardiac Imaging Associate Professor of Radiology New York University Medical Center	Research funding from Siemens
John Lesser, MD, FACC Director, Cardiovascular CT and MRI Minneapolis Heart Institute	Consulting relationships with Siemens and Vital Software
Robert McDonough, MD Senior Medical Director, Clinical Research and Policy Development Aetna, Inc.	Chair of pharmacy committee for Aetna; reviews technology for clinical research and policy group
James Min, MD Assistant Professor of Medicine, Division of Cardiology Assistant Professor of Radiology Weill Cornell Medical College & New York Presbyterian Hospitals	Not present at meeting

Peter J. Neumann, ScD Director, Center for the Evaluation of Value and Risk in Health, Institute for Clinical Research & Health Policy Studies Professor of Medicine Tufts-New England Medical Center & Tufts University	Consulting with GE Healthcare on project to develop metrics to value diagnostic technology
Mark Pauly, PhD Professor & Chair, Health Care Systems Wharton School University of Pennsylvania	Member of board of directors of non-profit payer
Rita Redberg, MD, MSc, FACC Director, Women's Cardiovascular Services Professor of Clinical Medicine University of California at San Francisco Medical Center	None
Donald Rucker, MD Vice President & Chief Medical Officer Siemens Medical Solutions USA	Employed by manufacturer
Sean Sullivan, PhD Director, Outcomes, Clinical Epidemiology, & Health Services Research Division Professor of Pharmacy Professor of Public Health/Community Medicine University of Washington	Pharmaceutical Outcomes Research and Policy Program (PORPP) receives funding from GE Healthcare for technology policy research
Sean Tunis, MD, MSc Founder & Director Center for Medical Technology Policy	None

INSTITUTE FOR CLINICAL AND ECONOMIC REVIEW

APPRAISAL OVERVIEW

CORONARY COMPUTED TOMOGRAPHIC ANGIOGRAPHY FOR DETECTION OF CORONARY ARTERY DISEASE

The overview is written by members of ICER's research team. The overview summarizes the evidence and views that have been considered by ICER and highlights key issues and uncertainties.

© 2009, Institute for Clinical and Economic Review

Final Scope

Rationale for the Appraisal
Coronary computed tomographic angiography (CCTA) is a minimally invasive radiological technique used to provide images of the heart and surrounding vessels. CCTA has been suggested as an alternative or useful complementary approach to other non-invasive methods of diagnosing coronary artery disease (CAD). In particular, because of its ability to visualize coronary anatomy, CCTA has been suggested as a strategy to rule out significant CAD among patients at low or intermediate risk of significant disease, thereby giving greater reassurance than other non-invasive methods and potentially reducing the number of patients ultimately sent for invasive coronary angiography (ICA). However, uncertainty remains regarding several important issues:

1) The diagnostic accuracy of CCTA relative to ICA and other possible comparator diagnostic tests
2) The impact on patient outcomes and health care utilization of alternative diagnostic algorithms that integrate CCTA in different ways into the diagnostic pathways for patients with suspected coronary artery disease, both in the general outpatient setting and in the Emergency Department
3) The most appropriate target populations for CCTA, based on level of risk and symptoms
4) The potential negative impact of increased radiation exposure of CCTA
5) The impact of incidental findings that trigger further evaluation
6) The potential impact of CCTA on the thresholds for clinician testing for coronary artery disease among the general population
7) The budget impact and cost-effectiveness of integrating CCTA into diagnostic pathways for patients with suspected coronary artery disease

Given the possible benefits of introducing a widely available non-invasive option for CAD detection, the potential clinical and financial impact that broad adoption of CCTA would have on systems of care, and the uncertainty over the evidence on the net health benefits and appropriate use of CCTA, all health care decision makers will benefit from a formal appraisal of the comparative clinical effectiveness and comparative value of CCTA as a modality for diagnosis of coronary artery disease.

Objective:
To appraise the comparative clinical effectiveness and comparative value of CCTA relative to the most relevant existing or emerging methods of CAD diagnosis and prognosis.

Key questions:

1. What are the sensitivity, specificity, and other test characteristics of CCTA in comparison to invasive coronary angiography as a reference standard but also in context with other accepted non-invasive modalities for CAD detection?

2. What is the impact of CCTA on diagnostic and treatment decision-making among patients being evaluated for possible coronary artery disease?

3. What is known about the impact of CCTA on patient outcomes?

4. How do CCTA's test characteristics vary according to important patient subgroups, such as gender and perceived risk or pretest probability of CAD?

5. What evidence exists on the frequency and outcomes related to incidental findings with CCTA?

6. What is known about CCTA's possible harms, including radiation exposure and contrast reactions?

Key considerations highlighted by the Evidence Review Group:

1. Target Population: While there has been some talk of CCTA's use as a screening tool in an asymptomatic population, current clinical opinion favors the use of CCTA only within a target population of symptomatic patients with low-to-intermediate likelihood of CAD. Insurers and clinical experts believe that an assessment of CCTA use within this patient population would yield the most important results for decision-making.

2. Setting: The two most relevant scenarios for use of CCTA include its use in (a) an ED setting for evaluation of acute chest pain; and (b) outpatient presentation with stable chest pain symptoms. CT calcium scoring for risk evaluation should not be considered by ICER at this time, as the major question among clinicians and payers has been focused on the use of CCTA to identify or exclude significant CAD.

3. Outcomes: While test performance is important to consider, emphasis should be given to consideration of evidence regarding CCTA's impact on diagnosis, therapeutic action, and patient outcomes. Within the literature on test performance, focus should be on "per-patient" findings rather than "per-vessel" or "per-segment", as clinical determination of CCTA interpretability in practice is made at the patient level.

4. Harms: Because other diagnostic tests used in combination with or instead of CCTA may also involve radiation, the total radiation dose of various diagnostic strategies should be considered. The fact that women often receive a higher dose of radiation should be noted. Also, new dose-reduction protocols should be considered within the body of evidence on CCTA radiation dose.

5. Ethical considerations: There appear to be no distinctive ethical issues regarding the patient population or the interpretation of results from cost-effectiveness analyses.

1. Background

1.1 The Condition

Coronary artery disease (CAD) is the leading cause of death in the United States among both men and women, resulting in over 400,000 deaths annually (Centers for Disease Control and Prevention and American Heart Association, 2008). CAD also has a substantial impact on health care utilization. For example, approximately 6 million patients are seen each year at emergency departments for acute chest pain, the hallmark symptom of CAD (Gallagher, 2007). Greater than 60% of hospitalizations for chest pain, costing more than $8 billion annually, are ultimately deemed unnecessary (Hoffmann, 2006).

CAD is caused by plaque accumulation and hardening in the coronary arteries, known as atherosclerosis. As buildup increases, the passage through the arteries narrows, decreasing blood flow and oxygen supply to the myocardium and causing angina and shortness of breath in many patients. Occlusion, or total blockage, of the arteries may result in myocardial infarction (Mayo Foundation for Medical Education and Research, 2008).

Due to its prevalence, and because several options (e.g., surgery, medication) exist to reduce CAD-related morbidity and mortality, accurate diagnosis of CAD is critical. Currently the definitive standard for diagnosis is invasive coronary angiography (ICA). There are risks associated with ICA, however, such as infection, artery trauma, and heart arrhythmias. For this reason non-invasive diagnostic methods have also been sought; the most common of these are the electrocardiogram (EKG), which measures cardiac activity via electrical signals, the echocardiogram (ECHO), which uses ultrasound to examine cardiac function, and single photon emission computed tomography (SPECT), which identifies abnormalities in cardiac perfusion using a radioactive tracer.

These tests differ in terms of their diagnostic accuracy, and their relative advantages and disadvantages. Because each test provides unique data, they are often used in combination when initial results are inconclusive. Given that none of the above-described tests provide a direct visual image of underlying coronary anatomy and degree of occlusion, interest has grown in using CT or MRI technology to evaluate patients with suspected CAD. Recently, the evolution of ultra-fast CT scanners has led to improved coronary imagery. Consequently, CCTA has received the endorsement of several clinical specialty organizations and is covered by many Medicare contractors and private insurers. Questions remain, however, regarding the relevant target populations for CCTA, its use alone or in combination with other tests, its prognostic ability, and its relative benefits and harms.

2. The Technology and its Comparators

2.1 Coronary CT Angiography

CCTA is a technique in which a CT scanner is used to acquire multiple simultaneous tomographic sections ("slices") of the coronary arteries. At the time of this outpatient procedure, an IV is placed into a peripheral vein and a contrast dye is administered for the purposes of visually defining the arteries for the scan. Beta blockers may be given to the patient to slow the heart rate in order to prevent artifacts of heart motion that may affect image quality. The patient is positioned on the CT scanner and a large number of x-ray images are taken from multiple angles and reconstructed using computer software. Multi-detector row CT scanners contain rotating gantries that capture multiple images, or "slices". A 64-slice CCTA was introduced in 2004 and increased the number of captured images from the previous 16- and 32-slice technology.

Improved spatial and temporal resolution from 64-slice machines has been found to shorten the time required to capture an image, decreasing motion artifact as well as reducing the time to conduct the entire scan to approximately 8 seconds (Mowatt, 2008). An advantage to the shorter scan time is that patient breath-hold requirements are lower, which in turn reduces the dose of contrast media and focuses enhancement primarily on cardiac structures (Leschka, 2005). In addition, improved resolution allows scanners to accommodate more patients with fast or irregular heart rates than was previously possible (Ratib, 2008).

The 64-slice scanner has rapidly replaced earlier versions and is currently considered to be the community standard for CCTA. In 2007, 256- and 320-slice CT scanners became available, but it is unclear whether the greater resolution of these versions will provide clinically relevant advances to 64-slice machines. Dual source 64-slice scanners have also been introduced in which two scanners are mounted on the gantry at 90 degree angles (Matt, 2007). Dual source scanning is claimed by some to further decrease procedure time, reduce heart motion artifacts, and lower the effective radiation dose to the patient.

In the emergency department, CCTA can be used for the triage of patients experiencing acute chest pain to "rule out" CAD as the underlying cause. In comparison to standard triage care, which involves the use of serial cardiac enzyme testing as well as stress testing where warranted, some commentators have postulated that CCTA may rapidly identify patients without underlying CAD, thereby reducing the number of patients referred for ICA and the observation time required by many patients awaiting less precise evaluation.

In the outpatient setting, CCTA is most often used to evaluate patients with stable, non-emergent symptoms. For such patients CCTA can be used as an initial test or as a method for further evaluation following inconclusive results from another non-invasive functional test. As is the case among patients in the ED, CCTA's possible advantages in the outpatient setting include the ability to visualize and quantify underlying CAD, which may allow for greater precision in determining subsequent treatment (e.g., angioplasty, bypass surgery, or medical management).

Compared to other non-invasive diagnostic methods there are also potential disadvantages specific to CCTA, including a small risk of allergic reaction from the use of contrast dye and the risk of renal damage from the dye among patients with pre-existing renal dysfunction. In addition, the increased precision from multi-detector row CT scanners is accompanied by a higher radiation dose to the patient. A number of protocols (e.g., prospective EKG gating, step-and-shoot methods) have been employed with varying degrees of success to reduce the radiation dose to the patient, but concern remains regarding the potential for increased risks of secondary malignancy.

One concern regarding CCTA is the impact of calcium accumulation in the arteries on its performance. It has been shown that high calcium scores, generally defined as Agatston scores higher than 400, lower the specificity of CCTA; patients with these high levels of calcification may comprise as much as one-quarter to one-third of candidates for CCTA (Raff, 2005; Mollet, 2005). The presence of high calcification in arteries has been cited as the primary cause of false positives in CCTA scans (Hoffmann, 2006). However, other studies have shown no effects of calcium score on diagnostic accuracy, but have found effects on the rate of non-diagnostic exams (Ho, 2008; Stoltzmann, 2008).

The point at which high calcium score negatively affects CCTA results is not universally agreed upon. Aetna, which covers CCTA for detection of CAD, considers CCTA to be investigational for patients with Agatston scores greater than 1700. Others have shown a threshold effect at an Agatston score of 600 (Miller, 2008).

As with any evolving technology, the expansion from academic centers to community practice may lead to variable competency in the interpretation of CCTA scans. Although the process of standardization of training in the conduct and interpretation of CCTA is underway (see Section 3), there is the possibility that the reported accuracy of CCTA could decrease as utilization by community practitioners, rather than clinicians at highly-specialized centers, increases.

Finally, the range of visualization of CCTA may extend beyond the heart itself, creating the possibility of identification of "incidental findings" that may or may not be related to the patients' complaints of chest discomfort. The clinical impact of incidental findings is controversial and will be the subject of subsequent discussion within this report.

2.2 Coronary Artery Disease Diagnosis Alternatives

For many years the most precise and definitive method for the evaluation and diagnosis of coronary artery disease has been invasive coronary angiography (ICA). At the time of the procedure a catheter is inserted into an artery, usually the femoral blood vessel, and contrast dye is injected through the catheter. X-ray images are then captured and displayed on a video screen (a procedure known as fluoroscopy), and can be viewed either as images or in motion picture form. While complications from ICA are relatively infrequent, they can be significant, and include myocardial infarction, cardiac arrhythmia, stroke, hemorrhage, infection, trauma to the artery from hematoma or from the catheter, sudden hypotension, and reaction to the contrast medium (Gandelman, 2006). The procedure also delivers a radiation dose lower than most CCTAs but similar to that of CCTA when it is performed using dose-saving protocols or dual-source scanners.

In part because of the invasive nature of ICA and its concordant risks, alternative non-invasive tests also are utilized for evaluation of chest pain symptoms considered suggestive of CAD. The first of these technologies to gain widespread use was the stress electrocardiogram (EKG); the major alternatives are stress echocardiography and single-photon emission computed tomography (SPECT), also known as nuclear stress testing or myocardial perfusion imaging.

Stress echocardiograms (ECHO) produce images of the heart through the use of sound waves. The test allows for the evaluation of blood flow in different areas of the heart to identify weak or damaged areas of the muscle. This is done through a comparison of images at rest and under cardiac stress induced by exercise or pharmacologic means. Clinically, the test is simple to perform, relatively inexpensive, and easily accessible. However, the image quality is lower in obese patients and those with chronic disease, which can account for almost 30% of candidates (Miller, 2006). It is recommended for use in intermediate-to-high risk patients (Anthony, 2005).

SPECT imaging involves the use of a tracer radiopharmaceutical to highlight areas of decreased blood flow in the myocardium. Images are captured via a gamma camera, and may be reconstructed to create two or three-dimensional films. The accuracy of SPECT imaging has improved to the point that it is often used for prognostic use in addition to diagnosis. However, it is not as effective in detecting perfusion defects in patient with milder stenosis (Jeetley, 2006). SPECT also involves the use of contrast media and delivers a radiation dose similar in magnitude to that of ICA and CCTA.

All of these alternative non-invasive diagnostic techniques measure in some way the functional impact on the heart of any underlying CAD. As noted above, none of the tests is perfect; each has the possibility of producing false positive and false negative results. Professional guidelines recognize all of these comparator techniques as appropriate initial investigations to evaluate possible CAD for most patients with stable symptoms (Gibbons, 2003).

3. Clinical Guidelines & Competency Standards

Published clinical guidelines on the use of CCTA are summarized here and presented in more detail in Appendix A.

- American Heart Association (2006)
  http://circ.ahajournals.org/cgi/reprint/CIRCULATIONAHA.106.178458

  CCTA has been shown to have a high negative predictive value, and therefore is useful in ruling out CAD. Evidence supports the use of CCTA for patients with low-to-intermediate probability of hemodynamically relevant stenosis and may obviate the need for ICA in these patients.

- Multi-Society Statement of Appropriateness Criteria for Cardiac Computed Tomography (2006) http://content.onlinejacc.org/cgi/content/full/48/7/1475

  Appropriateness reviews of CCTA and cardiac magnetic resonance imaging deemed the use of CCTA for detection of CAD to be appropriate for the following patient populations:
  - Presenting with chest pain syndrome with intermediate pre-test probability of CAD and uninterpretable EKG or inability to exercise
  - Presenting with chest pain and uninterpretable or equivocal stress test results
  - Presenting with acute chest pain with intermediate pre-test probability of CAD and no EKG changes and serial enzymes negative
  - Symptomatic patients requiring evaluation of suspected coronary anomalies

- American College of Radiology (2006)
  http://www.acr.org/SecondaryMainMenuCategories/quality_safety/app_criteria/pdf/ExpertPanelonCardiovascularImaging/ChronicChestPainSuspectedCardiacOriginUpdateinProgressDoc8.aspx

  An update to their 1995 recommendations determined that CCTA is appropriate for assessment of CAD, although its usefulness for patients with low pretest probability is unknown. On a scale of 9 to indicate appropriateness (with a score of 9 being most appropriate), CCTA was assigned a rating of 7 for the evaluation of chronic chest pain.

- SCCT/NASCI Consensus Update (2007)
  http://www.invasivecardiology.com/article/7959

  An update to their 2006 publication found CCTA to be appropriate in the following circumstances:
  - To rule out significant coronary stenosis
  - To evaluate patients with equivocal or discordant results on a stress perfusion or wall motion study

- To rule out stenosis in patients with a low pre-test likelihood of CAD
- To potentially replace diagnostic catheterization in patients undergoing non-coronary cardiac surgery

- ACCF/AHA Clinical Competence Statement (2005, updated 2007)
  http://www.scct.org/ct_mr_clinical_competence_statement_063005.pdf

  Guidelines for the assessment of clinical competence of physicians performing CCTA were established. The minimum training required to independently perform and interpret CCTA, both non-contrast and contrast, is as follows:
  - Board certification of eligibility and valid medical license
  - Eight weeks of specialized training in CCTA
  - 150 contrast CCTA examinations (at least 50 in-person)
  - Evaluation of 50 non-contrast studies
  - Completion of at least 20 hours of courses related to general CT or CCTA

- ACR Practice Guidelines (2006)
  http://www.acr.org/SecondaryMainMenuCategories/quality_safety/guidelines/dx/cardio/ct_cardiac.aspx

  Physician competency in performing and interpreting CCTA is defined by the following qualifications:
  - For physicians with prior qualifications for interpretation of CT examinations, a minimum of 30 hours of training courses in cardiac anatomy, physiology, and pathology and at least 50 CCTA examinations supervised, interpreted, or reported in the last three years
  - For physicians with no prior qualification, a minimum of 200 hours of training on performance and interpretation of CT and supervision, interpretation, and reporting of at least 500 cases (at least 100 must be thoracic CT or CCTA), in addition to the training and interpretation requirements specified above
  - Understanding of administration, contraindications, and risks of pharmacologic agents used for CCTA
  - Continuous use of the technology, defined as a minimum of 75 cases per three years
  - Continuing medical education relevant to CCTA

NOTE: There is now a formal board certification process for cardiologists wishing to be certified in cardiac CT imaging that is being administered on behalf of multiple clinical societies (ACC, ASNC, SCAI, and SCCT). Candidates must meet minimum ACCF/AHA criteria, undertake a formal examination, and be re-certified every 10 years. (http://www.cbcct.org/index.cfm) A similar effort is being undertaken by the ACR on behalf of radiologists.

4. Medicare and Representative Private Insurer Coverage Policies

- In December 2007, citing CCTA as a promising but unproven technology, the Centers for Medicare and Medicaid Services (CMS) announced its intent to create a national coverage decision (NCD) allowing for "coverage with evidence development"—that is, coverage only for patients participating in clinical trials of the technology. After a period of public comment and discussion, CMS reversed its decision in March 2008, and stated that the local coverage determination (LCD) process would be left in place. Current LCDs allow for coverage of CCTA in symptomatic patients, but some recent LCDs place additional restrictions on coverage; below is an example of the covered indications from the regional contractor for the states of Alaska, Oregon, and Washington:

    - In the emergency room setting, CCTA is covered for evaluation of patients with acute chest pain or for first-line testing for CAD among intermediate risk patients.
    - CCTA is allowed for ruling out CAD among low-to-intermediate risk patients following equivocal test results where negative results will avoid invasive coronary angiography.
    - CCTA is covered for assessment of surgical eligibility among patients with congenital anomalies of the coronary vessels or greater vessels or for patients in sinus rhythm scheduled for non-coronary cardiovascular surgery.
    - CCTA is covered for evaluation of pulmonary veins and atrium in patients with atrial fibrillation and/or flutter when evaluation avoids what would otherwise be a medically reasonable and necessary MRI in patients who are scheduled to undergo ablation therapy evaluation.
    - CCTA is *not* covered in any other circumstance, including for screening, demonstration of coronary calcification, or risk stratification.

- Among private health plans with publicly available coverage policies for 64-slice CCTA, details of coverage differ. Representative examples of coverage policies include the following:

    - Aetna covers 64-slice CCTA for ruling out CAD in patients with low pre-test probability and equivocal or contraindicated stress testing, conducting pre-operative assessments for non-coronary cardiac surgery, detection of coronary anomalies, evaluating cardiac structures in patients with congenital heart disease, and calcium scoring.

    - CIGNA covers 64-slice CCTA for detection of CAD in symptomatic patients with intermediate pre-test probability and equivocal or contraindicated EKG, or with no EKG changes and negative enzymes.

- United Healthcare covers 64-slice or better CCTA for evaluation of chest pain among patients with intermediate pre-test CAD probability and equivocal or contraindicated EKG, evaluation of chest pain among patients with prior uninterpretable or equivocal stress test results, and assessment of acute chest pain in patients with an intermediate pre-test probability of CAD, no EKG changes, and negative enzymes.

- The Regence Group and UniCare both consider CCTA to be investigational and will cover its use only if ICA was unsuccessful or equivocal for detection of CAD.

5. Previous Systematic Reviews/Tech Assessments

- U.K. National Health Service Research & Development Health Technology Assessment (2008)
  http://www.ncchta.org/execsumm/summ1217.shtml
  CCTA will most likely not replace ICA, but may be useful in ruling out significant CAD.

- BCBSA TEC (2006)
  http://www.bcbs.com/blueresources/tec/vols/21/21_05.html
  Evidence on CCTA for use in either diagnosis of coronary artery stenosis or evaluation of acute chest pain does not meet TEC criteria for widespread adoption and use. The only criterion that was met was the first, which states that "the technology must have final approval from appropriate government regulatory bodies". The following criteria were not met:
    - The scientific evidence must permit conclusions concerning the effectiveness of the technology regarding health outcomes.
    - The technology must improve net health outcomes.
    - The technology must be as beneficial as any of the established alternatives.
    - The improvement must be attainable outside the investigational settings.

- Medical Services Advisory Committee (MSAC) (2007)
  http://www.msac.gov.au/internet/msac/publishing.nsf/Content/app1105-1
  In symptomatic patients, CCTA is as effective as ICA in ruling out significant CAD.

- California Technology Assessment Forum (CTAF) (2007)
  http://www.ctaf.org/content/general/detail/768
  CCTA for diagnosis of coronary artery stenosis and evaluation of acute chest pain failed to meet CTAF criteria for widespread adoption and use. Criteria utilized by CTAF were the same as those of BCBSA TEC; the only criterion that was met was Criterion 1, which states that "the technology must have final approval from appropriate government regulatory bodies".

- Medicare Coverage Advisory Committee (MedCAC) (2006)
  http://www.cms.hhs.gov/mcd/viewmcac.asp?where=index&mid=34
  While individual responses varied, the committee's response was "unsure" when questioned as to whether 64-slice CCTA would provide a net health benefit when (a) used as a non-invasive diagnostic test before ICA; or (b) used as a replacement for ICA.

- Ontario Health Technology Advisory Committee (2005)
  http://www.health.gov.on.ca/english/providers/program/ohtac/tech/reviews/sum_mdct_20070926.html
  There is insufficient evidence to suggest that 16- or 64-slice CCTA is equal to or better than coronary angiography to diagnose CAD in those with symptoms or to monitor progression in persons with prior cardiac interventions.

- National Institute for Health and Clinical Excellence (NICE)
  http://www.nice.org.uk/guidance/index.jsp?action=byID&o=11953
  NICE has not reviewed this topic.

6. Ongoing Clinical Studies

Thirty clinical studies are currently recruiting patients for evaluation of CCTA as a diagnostic tool for CAD; four are randomized studies and two are employing within-subject designs to compare CCTA with ICA or SPECT. Several large cohort studies are documenting CCTA in clinical practice. Major studies are summarized below (details at http://clinicaltrials.gov).

Table 1. Summary of ongoing clinical studies

Trial Sponsor	Design	Primary Outcomes	Populations	Variables	Comments
Beaumont Hospitals (NCT00541203)	RCT	▪ Diagnostic/prognostic performance ▪ Prediction of major cardiovascular events	N=200 with inconclusive or indeterminate stress test results	CCTA vs. ICA	Outpatient setting; Estimated study completion date January 2012
Seoul National University (NCT00431977)	RCT	▪ Myocardial infarction ▪ Late revascularization ▪ Cardiac death	N=1,000 diabetics without coronary symptoms	CCTA+ standard care vs. standard care	Estimated study completion date December 2012
Intermountain Healthcare (NCT00488033)	RCT	▪ All-cause death ▪ Non-fatal MI ▪ Unstable angina	N=1,100 asymptomatic, high-risk diabetics	Screening with CCTA or calcium scoring vs. standard care	Estimated study completion date December 2011
Beaumont Hospitals (NCT00468325)	RCT	▪ Multiple efficacy, safety, and economic endpoints	N=750 ED patients with acute chest pain and low-to-intermediate CAD risk	CCTA vs. standard triage care	Emergency Department setting; Estimated study completion date December 2008
St. Joseph's Healthcare (NCT00371891)	Within-subject	▪ Sensitivity and specificity	N=900 scheduled for ICA	CCTA vs. ICA, single-blinded comparison	
GE Healthcare (NCT00486447)	Within-subject	▪ Sensitivity, specificity ▪ Negative predictive value ▪ Downstream cardiac testing ▪ Major cardiac events	N=300 with intermediate CAD risk and referred for myocardial perfusion scanning	CCTA vs. MPS, single-blinded comparison	Estimated study completion date August 2011

Table 1. Summary of ongoing clinical studies (cont'd)

Trial Sponsor	Design	Primary Outcomes	Populations	Variables	Comments
Brigham and Women's Hospital (NCT00321399)	Observational	▪ Referral to cardiac catheterization within 90 days of index test ▪ Predictive ability for cardiac death and non-fatal myocardial infarction ▪ Relative cost-effectiveness of each approach	N=4,000 referred for stress perfusion (SPECT, PET), CCTA, or combined perfusion-anatomy (PET/CT) studies with intermediate-to-high pretest probability of CAD	CCTA vs. PET, SPECT, and hybrid PET-CT	Estimated study completion date August 2009
William Beaumont Hospitals (NCT00640068)	Observational	▪ Patient characteristics ▪ Scanning acquisition techniques ▪ Quality of physician scan interpretation ▪ 90-day clinical outcomes	N=12,000 referred or self-referred for CCTA	CCTA	Study was a collaborative effort organized by Blue Cross/Blue Shield of Michigan; estimated study completion date October 2010

7. The Evidence

7.1 Systematic Literature Review

Objectives
The primary objective of the systematic review was to identify and summarize the published evidence on the test performance and impact on patient outcomes of CCTA in two key populations:

- Acute chest pain of unknown origin in an ED setting
- Stable chest pain symptoms among patients at low-to-intermediate CAD risk in an outpatient setting

We sought studies that prospectively examined the impact of CCTA, whether used alone or in combination with other diagnostic methods, on objective outcomes; these included treatment and testing decisions and major cardiovascular events. We also included studies that evaluated CCTA's diagnostic accuracy relative to a common reference standard (typically ICA). While we did not systematically search for evidence regarding test safety, incidental findings, and economic impact, we obtained such data within our selected clinical literature, supplemented with data from review articles and expert guidance.

Many candidate studies reported results on a "per-vessel" or "per-segment" basis, in addition to per-patient analyses. While these approaches are often useful for juxtaposing segment or vessel location against temporal and spatial resolution on CCTA, and provide a larger sample of observations in which to examine accuracy, they are not generalizable to clinical practice, in which decisions on patient management are made at the patient level. For example, a distal segment may be excluded from analyses of accuracy because of blurred imagery; in reality, any indeterminate finding on any segment can trigger further testing at the patient level. Because of our interest in examining the impact of CCTA on patient outcomes, and because per-vessel results alone can inflate test performance statistics, we included only those studies that reported results at the patient level or whose results could be used to construct per-patient analyses.

Methods
This review included studies of the performance of CCTA in diagnosing CAD using scanners with 64-slice or higher resolution. Guidance from the ICER Evidence Review Group suggested that 64-slice scanners were now widely available in the community and had become viewed as the standard for CCTA, and that literature on earlier-generation scanners would not be viewed as relevant by the clinical and patient communities.

We also excluded studies that reported on the use of CCTA for applications other than CAD detection—for example, diagnosis of pulmonary emboli or detection of congenital cardiac defects. We also excluded studies focused solely on the use of CT for so-called "calcium scoring", or measurement of coronary calcium as a marker for early-stage CAD, as

the focus of our appraisal was on the diagnosis of obstructive disease among symptomatic patients.

Included studies were conducted in ED or outpatient settings (as described above) and had a study population of adults who underwent CCTA. Studies of diagnostic accuracy must have used ICA as the reference standard in all or a random sample of patients. We searched for studies during the period January 2005 (the first year of published studies from 64-slice scanners) to the present. Other major eligibility criteria included:

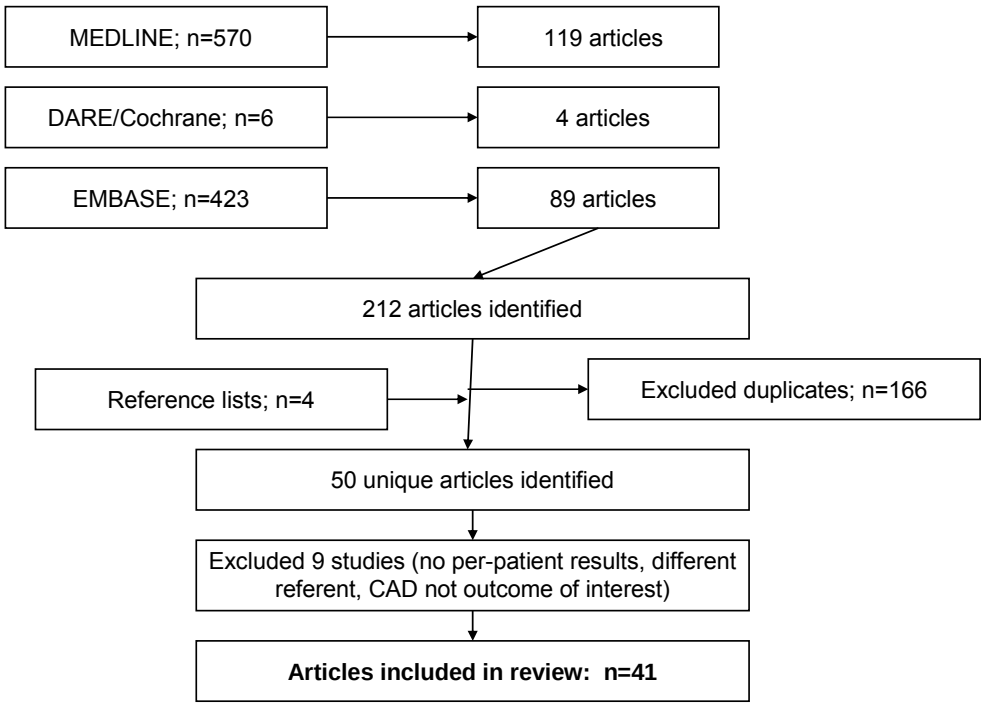
- Results reported on per-patient basis (or ability to construct per-patient findings)
- Receipt of reference standard by entire study population or random sample
- For diagnostic accuracy studies, time between CCTA and reference standard did not exceed 3 months
- Evaluation of native arteries only
- Blinded review of both CCTA and reference test

Studies were not further restricted by CCTA instrumentation, imaging technology, method of heart rate control, or use and type of dose-sparing protocol.

Electronic databases searched included MEDLINE, EMBASE, and *The Cochrane Library* (including the Database of Abstracts of Reviews of Effects [DARE]) for eligible studies, including health technology assessments (HTAs), systematic reviews, and primary studies. Reference lists of all eligible studies were also searched. The search strategies used for MEDLINE, EMBASE, and *The Cochrane Library* are shown in Appendix B.

On the following page, Figure 1 shows a flow chart of the results of all searches for included primary studies. In addition to 41 primary studies, searches identified 2 systematic reviews and 2 HTAs.

Figure 1. QUORUM flow chart showing results of literature search
Data abstracted from each primary study included inclusion and exclusion criteria, patient demographics and risk status (if available), sample size, # of patients with known prior CAD, # of patients excluded for non-diagnostic CCTA results, stenosis threshold for CAD diagnosis, sensitivity, specificity, PPV, and NPV for significant CAD (by patient only), prevalence of CAD by number of diseased vessels (based on reference standard), complications, and effective radiation dose.



7.2 Data Analyses

Patient Outcomes Data

Because studies of the impact of CCTA on clinical outcomes varied in terms of their definitions of events, period of follow-up, and data collection methods, we made no attempt to formally meta-analyze these data. Study characteristics and major findings are presented in descriptive fashion only, and general trends and/or consistencies across the studies are discussed.

Diagnostic Accuracy Data

If sensitivity or specificity was not reported, we calculated these values. We calculated sensitivities whenever true positive and false negatives values were reported using the formula "true positive/ (true positive + false negative)"; negative predictive value (NPV) was calculated as "true negative/(true negative + false negative)". Specificity was calculated using the formula "true negative/(false positive + true negative)", and positive predictive value (PPV) as "true positive/(true positive + false positive)".

We present published data according to an "intent to diagnose" (ITD) paradigm; in this approach, patients with "non-diagnostic" or indeterminate CCTA tests are considered to have positive findings, as clinical expert guidance from the ICER Evidence Review Group suggested that clinicians commonly refer such cases to ICA or further non-invasive testing.

Our primary approach conservatively assumed that all such patients would be determined to be false positives on ICA, which materially affects only the calculations of specificity and PPV (i.e., as false positives are not included in calculations of sensitivity or NPV). This approach may under-represent the diagnostic accuracy of CCTA but avoids the equal or greater risk of overestimating accuracy when non-diagnostic CCTA results are excluded from consideration. This "conservative" approach has been employed by several investigators (Ropers U, 2007; Shapiro, 2007) specifically to evaluate the impact of excluding non-diagnostic findings on test characteristics.

The quality of diagnostic accuracy studies is typically assessed using the QUADAS tool, a 14-item instrument evaluating internal validity developed by Whiting et. al (Whiting, 2003). We modified the published tool by first eliminating 2 items that relate to sufficient description of the index test and reference standard to allow their replication, as it was felt that these items relate more to the quality of study reporting rather than any methodological deficiencies. We then added 4 items to the checklist, consistent with methods used in a recent HTA and systematic review of 64-slice CCTA (Mowatt, 2008):

- Use of an established threshold to define stenosis
- Presentation of data on inter-observer variation and results within acceptable ranges
- Data presented for appropriate patient subgroups
- Reporting of true disease prevalence on ICA (or ability to derive it)

The modified QUADAS tool is presented in Appendix C, along with the results of our study quality review.

Data Synthesis

Analyses of test characteristics were conducted by first using the reported or derived numbers of true positives, false positives, true negatives, and false negatives to calculate sensitivity, specificity, PPV, and NPV. These statistics were used in turn to generate the positive likelihood ratio (PLR, increase in odds of disease with positive test result) and negative likelihood ratio (NLR, decrease in odds of disease with negative test result).

We generated summary receiver operating characteristic (sROC) curves to assess whether any threshold effects appeared to be present, and correspondingly, whether symmetric or asymmetric distributions should be assumed. Pooled estimates of test accuracy were generated using the DerSimonian-Laird method for random-effects models (DerSimonian, 1986); 95% confidence intervals were also constructed.

In addition to primary analyses of data, alternative analyses were conducted to: (a) examine the influence of inclusion of patients with known CAD in the study sample by comparing pooled results between studies that did and did not include such patients; and, (b) assess the effect of excluding patients with non-diagnostic results by comparing overall pooled results to findings recalculated using the ITD approach. Meta-analyses were conducted using MetaDiSc software version 1.4 (Zamora, 2006).

7.3 Results

Selected Studies

A total of 50 studies were initially identified from the literature search; 9 of these studies were excluded because either no per-patient findings were available (n=4), the comparison performed was for an outcome other than detection of CAD (e.g., comparison to SPECT to assess myocardial perfusion, n=2), or identical/overlapping study samples were presented in another included study (n=3). Characteristics of excluded studies are presented in Table 1 at the end of this report.

Of the remaining 41 studies, 33 were conducted in an outpatient setting, and 8 were conducted in an ED setting. Most studies were diagnostic accuracy studies using ICA alone as the reference standard (n=32; 1 ED, 31 outpatient), with most of these conducted in patients already scheduled for ICA. A total of 9 studies examined the impact of CCTA by evaluating subsequent clinical decisions and patient outcomes; while this approach was typically utilized in an ED setting (where definitive diagnosis by ICA is not universally feasible or warranted), 2 of the 9 studies identified were conducted among patients presenting on an outpatient basis with stable symptoms. Characteristics of included studies are presented in Table 2.

Because most of the included studies involved patients already scheduled for ICA, the prevalence of CAD in our sample was relatively high (mean [SD]: 59.0% [20.9%]; range: 18.2%-91.0%). Studies reporting results stratified by CAD risk or pretest likelihood are summarized below.

Major reasons for patient exclusion from these studies related primarily to ability to perform CCTA or obtain adequate image quality, and included known allergy to contrast media, impaired renal status, inability to follow breath-hold commands, obesity (typically, defined as BMI >40), and elevated heart rate after attempted pharmacologic control. Approximately two-thirds of studies also excluded patients with known prior CAD or revascularization. Finally, while not a criterion for *patient* exclusion, vessels smaller than 1.5 mm in diameter or those felt to be heavily calcified were often excluded from analysis, as CCTA image quality is often impaired in these vessel types (Schroeder, 2008).

All of the selected studies were conducted in single centers; findings from the first published multi-center study were available after our analyses were completed, but are summarized later in this Section under "Additional Recent Evidence". Two randomized studies were identified; a randomized controlled trial of standard ED triage care to CCTA plus standard care (Goldstein 2007), and a randomized comparison of dual-source to single-source CCTA (Achenbach 2008). Characterization of selected studies according to a widely-accepted framework for assessing the level of evidence from diagnostic imaging studies (Fryback 1991) can be found below (from lowest to highest level of evidence presented):

1. Technical only: 0
2. Diagnostic accuracy: 32

3. Impact on diagnostic thinking: 2
4. Impact on therapeutic actions: 6
5. Impact on patient outcomes: 1
6. Impact on societal outcomes: 0

Importantly, while there were 9 studies in our sample that measured outcomes beyond test accuracy, only the Goldstein study evaluated the incremental effects of CCTA relative to a comparison group, and was therefore the only study identified as measuring the attributable impact of CCTA on patient outcomes.

Description of Study Population

ED Studies

A total of 9 reports were initially identified that examined CCTA's impact on outcomes or diagnostic accuracy in the ED setting, one of which was excluded from the final sample. This study (Rubinshtein, 2007 [2]) was based on an identical sample reported in another publication that was included in our final sample.

The total sample size in the ED studies was 679 patients; sample size ranged from 33-104 by study. Mean age ranged from 46-58 years; approximately 60% of the overall sample was male. The presence of prior known CAD or ischemia was observed in about 7% of patients (n=34).

Outpatient Studies

A total of 41 reports were initially identified that examined CCTA diagnostic accuracy in the outpatient setting. Eight of these studies were excluded, because results were not reported on a per-patient basis (n=4) or ICA was not part of the reference standard definition (n=2), or the study sample overlapped with another from the same institution (n=2).

The total sample size in the remaining 33 studies was 3,559, and ranged between 30-421 patients per study. Mean age ranged between 46-69 years; 63% of the overall sample was male. The overall prevalence of prior known CAD was approximately 10%, and ranged between 2-40% in those studies including patients with known prior disease.

Studies of CCTA Impact on Clinical Decisions and Patient Outcomes

Details on the 9 studies that evaluated in some way the impact of CCTA on patient management and outcomes can be found in Table 3. The outcome measures employed, event definitions used, underlying CAD risk, and duration of follow-up varied significantly between studies. In addition, the lack of active or historical controls in all but one of these studies made CCTA's possible incremental benefits and health-system impacts difficult to ascertain. Brief descriptions and key findings of these studies are given below.

Goldstein (2007): This study was an RCT of CCTA plus standard triage care vs. standard care alone in 197 patients at very low risk of CAD. Following initial negative EKGs and serum enzymes for myocardial damage, patients in the CCTA arm were discharged home

immediately if they had a normal study or non-significant CAD, referred for ICA if CCTA indicated severe stenosis, or referred for standard triage care if CCTA results indicated intermediate stenosis or were non-diagnostic. Seventy-five percent of patients in the CCTA arm were discharged home immediately, and none of these patients suffered major cardiac events over a 6-month follow-up period. A higher percentage of patients in the CCTA arm had ICA; 9 of the 11 catheterizations in the CCTA arm confirmed significant CAD. One of 9 patients (11%) with a positive CCTA was determined to be a false positive on ICA. Testing costs were higher in the CCTA arm, but due to shorter average ED stays total ED costs per patient were approximately $300 lower for the CCTA arm.

Rubinshtein (2007) [3]: This study evaluated CCTA's use in guiding triage among 58 patients with and without known prior CAD who presented to the ED with chest pain, intermediate CAD risk, negative initial enzymes, and no EKG changes. Patients received standard ED triage along with cardiology consultation, after which a presumptive diagnosis of acute coronary syndrome (ACS) was made where warranted with recommendations for hospitalization and early invasive treatment. CCTA was then performed in all patients, and recommendations adjusted based on CCTA findings. Patients were followed for major adverse cardiovascular events (MACE) over a mean of 12 months of follow-up. CCTA results led to a revised ACS diagnosis in 18 of 41 patients, canceled hospitalizations in 21 of 47, and altered early invasive treatment in 25 of 58. One CCTA scan was deemed to be false positive; no MACE events were recorded in the 32 patients discharged from the ED.

Pundziute (2007): The prognostic significance of CCTA was evaluated in this study of 100 outpatients who were referred for further evaluation (stress EKG, SPECT, or ICA) based on suspicion of CAD. CCTA and calcium scoring were performed in addition to the standard workup. A total of 26 patients had at least one MACE event over a mean follow-up of 16 months. In Kaplan-Meier analyses of event rates at one year, a positive CCTA for any stenosis was associated with a significantly increased event risk (30% vs. 0%, p=.005); whether CAD was deemed to be obstructive on CCTA, as well as location of obstructive disease, were significant and independent predictors of event likelihood.

Hollander (2007): A total of 54 low-risk patients presenting to the ED with chest pain and negative initial enzymes were scheduled for EKG and CCTA in this study. The incidence of MACE events was recorded at 30 days post-ED visit. A total of 46 patients (85%) were immediately discharged from the ED after negative CCTA findings; no MACE events were recorded among these patients. Two of the remaining patients were hospitalized even though CCTA findings were negative (the ED physician did not yet have enough confidence in the technology); of the remaining 6 patients, 2 had high degrees of stenosis confirmed by ICA, and 4 were referred for subsequent non-invasive testing after moderate stenosis was observed on CCTA. No events were observed in any patient at 30 days.

Gallagher (2007): A total of 85 low-risk ED patients (7 were excluded due to uninterpretable CCTA scans) with suspected ACS received both stress SPECT and CCTA after admission to a chest pain observation unit; subsequent triage was based on the combined results of the

two tests. Patients were followed for MACE events at 30 days. The majority of patients (85-86%) had negative findings on either test. A total of 7 patients had confirmed ACS; among these patients, one had a negative CCTA and positive SPECT, 2 had a positive CCTA and negative SPECT, and 4 had a positive result on both tests. No events were recorded in any patient at 30 days.

Johnson (2007) [1]: In this study, 55 patients with acute chest pain of unknown origin were referred from the ED for CCTA and followed for at least 5 months for the cause of chest pain (both CAD and non-CAD) as well as long-term outcomes. CCTA identified the cause of chest pain in 37 of 55 patients (67%); in 14 patients, neither CCTA nor clinical follow-up determined the cause of chest pain; and in 4 patients, a diagnosis was made from clinical follow-up only.

Savino (2006): Early experience with CCTA was documented in this study of 23 patients presenting to the ED with acute chest pain and no EKG or enzyme changes. Short-term outcomes, including length and costs of hospitalization, were measured for study patients in comparison to a demographically-similar control group undergoing conventional ED workup. Of the 23 patients, 8 were identified as having ≥50% stenosis in at least one artery, which was confirmed by ICA in all cases; 2 were identified as having mild stenosis, received medical therapy and were discharged; 2 were identified as having pulmonary embolism, and were treated and discharged; and 11 were CCTA-negative (9 were immediately discharged). Length of stay and costs were reduced by ~40% in the study group relative to controls.

Danciu (2007): In this study, a total of 421 patients with symptoms suggestive of CAD and intermediate-risk results on stress SPECT were referred for CCTA. Patients with severe stenosis on CCTA or moderate stenosis that matched a perfusion deficit on SPECT received immediate ICA; those with moderate stenosis not matching a perfusion deficit, mild stenosis, or no stenosis on CCTA were medically managed. The majority of patients (81.5%) were medically managed based on combined SPECT-CCTA findings; among these patients, 6 (1.7%) had recurrent symptoms requiring late (>1 month) ICA, and one (0.3%) required late revascularization. The combined rate of death, MI, and any revascularization was 0.3% among medically-managed patients vs. 70.5% among those referred for ICA.

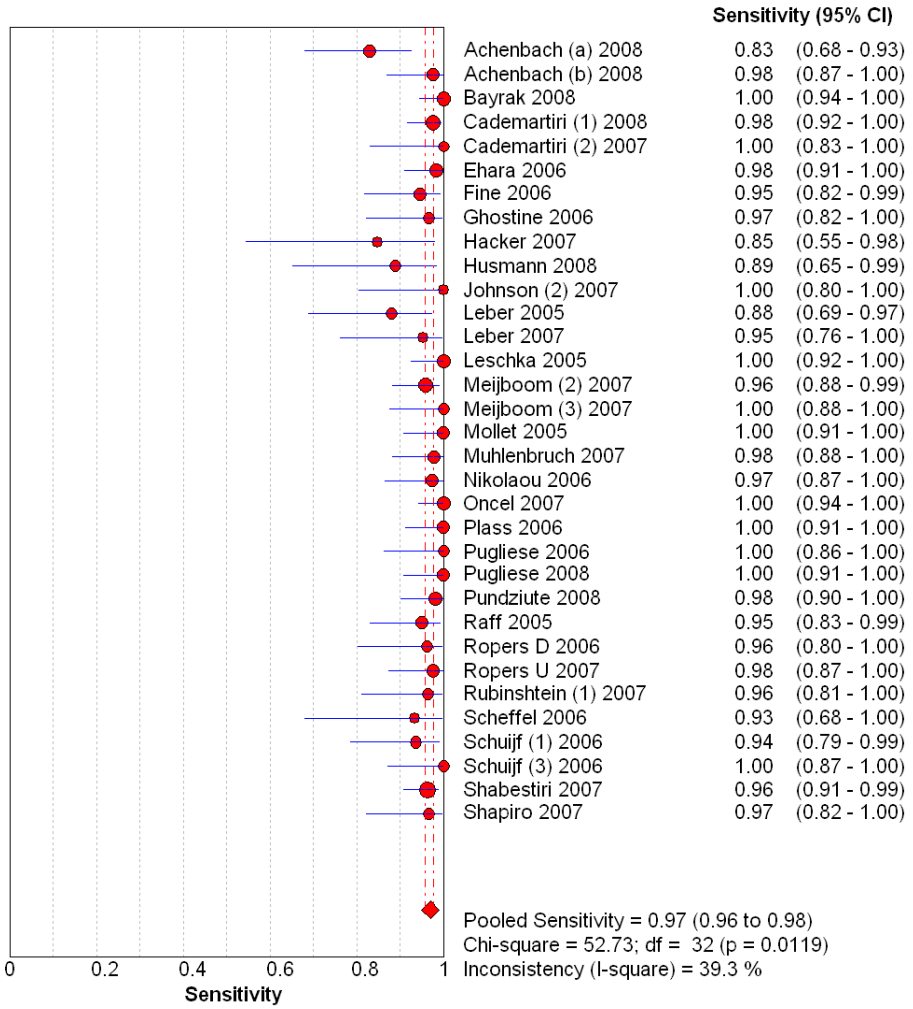
Hoffmann (2006). The potential effects of CCTA's identification of significant stenosis as well as calcified and non-calcified plaque were explored in this blinded prospective study of 103 patients presenting with acute chest pain, no EKG changes, and negative enzymes; all patients were hospitalized to rule out ACS. Patients were administered CCTA immediately prior to hospital admission. The presence of ACS was determined by an independent panel based on data collected during the index hospitalization and 5 months of follow-up. A total of 14 cases of ACS were identified; CCTA did not show evidence of significant stenosis in 73 patients (none of whom had ACS), detected significant stenosis in 13 patients (8 with ACS), and could not rule out stenosis in 17 patients (6 with ACS). Quantification of plaque by CCTA was an independent and significant predictor of ACS on

logistic regression analyses that included traditional risk factors (e.g., age, gender, hypertension).

CCTA Diagnostic Accuracy vs. ICA

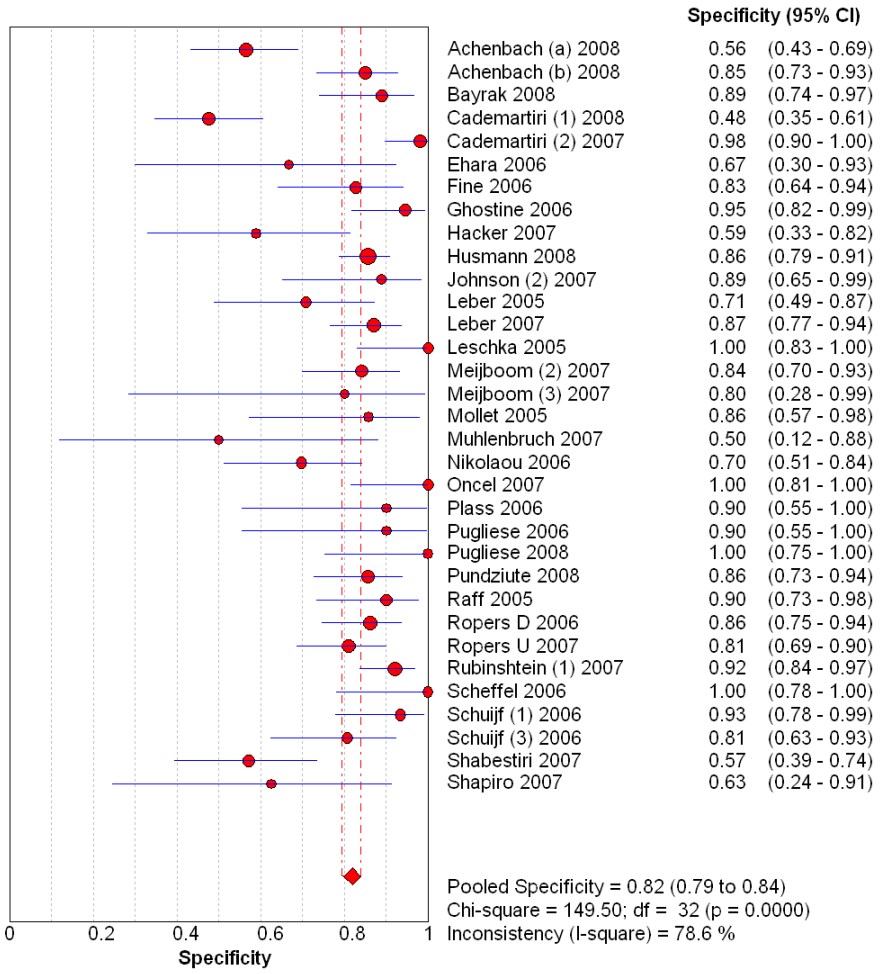
Figure 2 below presents the data on sensitivity of CCTA when compared directly to ICA, including the pooled results generated by quantitative meta-analysis. Where multiple subgroups (e.g., by CAD risk or gender) were reported, we considered these groups separately (yielding a total of 39 observations). The pooled sensitivity was 97% (95% CI, 96%, 98%); estimates were relatively consistent across studies (see Figure 2 below and Table 4 at the end of the report). Summary ROC curves (Appendix D) showed no evidence of a threshold effect, which was likely due to a relatively standard cutoff for identifying stenosis (≥50% luminal narrowing). About 3% of patients had non-diagnostic CCTA results (range: 0-18%); as described above, we included these patients as false positives in primary calculations.

Figure 2. Pooled sensitivity of CCTA in diagnosing CAD (intent-to-diagnose analysis).



A greater degree of variability was observed in analyses of specificity; results by study ranged from 50-100%. No discernible pattern in study design or diagnosis confirmation was observed among "outlier" studies. Consideration of patients with non-diagnostic findings as false positives resulted in a pooled specificity estimate of 82% (95% CI: 79%, 84%). Findings by study are displayed in Figure 3 below as well as in Table 4 at the end of this report.

Figure 3. Pooled specificity of CCTA in diagnosing CAD (intent-to-diagnose analysis).



NLR and PLR findings echoed those of sensitivity and specificity (Appendix D). When results from the diagnostic accuracy studies were pooled but with non-diagnostic exams excluded from consideration, specificity rose from 82% to 87% (95% CI: 85%, 89%). Full results for this alternative approach are shown in Appendix E. As discussed earlier, whereas our primary approach to determining specificity may under-represent CCTA performance, it is not unreasonable given the likelihood that excluding non-diagnostic exams ignores the fact that many patients with such results will be felt to require further

investigation, even though the true prevalence of significant disease among these patients is relatively low.

Given that CCTA has been a rapidly evolving technology, it is always possible that a pooling of evidence from studies published over several years will trail behind the most recent results. We examined this possibility but found that our estimates of sensitivity and specificity from pooling of studies 2005-2008 are similar to those from the most recent reports of CCTA diagnostic accuracy (i.e., Budoff, 2008; Bayrak, 2008; Husmann, 2008 [1]; Pundziute, 2008).

Formal Examination of Heterogeneity and Publication Bias

The I^2 statistic was generated for pooled estimates of sensitivity and specificity. For sensitivity, low-moderate heterogeneity was observed (I^2=39.3%; chi-square=52.73, p=.0119); however, a high degree of heterogeneity was seen in analyses of specificity (I^2=78.6%; chi-square=149.50, p=.0000). To further explore explanations for heterogeneity, the possibility of a "threshold effect"--i.e., variability in the cognitive threshold necessary for an investigator to call a patient diseased--was first examined. While an inverse relationship between sensitivity and specificity was seen (Spearman correlation coefficient: -0.169), this was not statistically significant (p=.333), indicating no material threshold effect. However, the absence of a significant threshold effect does not fully mitigate concerns regarding imprecise estimation of accuracy; for example the high underlying prevalence of CAD in many of these studies may raise concerns of "spectrum bias". In this case, failure to include lower-risk individuals might lead to over-estimation of diagnostic accuracy (Goehring, 2004).

A meta-regression model was then specified, including sample size, mean age, % male, and whether the sample included patients with known prior CAD. Only age was significant in the model; in further subgroup analyses focusing on the diagnostic odds ratio (a combined statistic incorporating sensitivity and specificity), heterogeneity increased substantially with increasing mean population age. However, pooled sensitivity and specificity in the youngest age group (i.e., the group with low heterogeneity) was quite similar to the overall as-reported results (98% and 87% respectively), so further controlling for age in the meta-analysis was not felt to be necessary. The results of heterogeneity analyses can be found in Appendix G.

Because the evidence base for this meta-analysis was limited to single-center studies, published in many cases by the same primary authors, publication bias was also considered as a partial explanation for the results. Authors with more than one publication in the sample were first contacted to identify whether any samples had a significant degree of overlap; in such cases, the largest study was retained and the others were removed. Authors were also queried regarding the presence of any significant unpublished or grey literature research that might have altered the primary findings; no studies were identified. Formal tests for publication bias were not conducted, as several recent analyses have suggested that most meta-analyses of small, single-center diagnostic accuracy studies are

subject to some level of publication bias (Song, 2002; Deeks, 2005). Uncertainty in pooled estimates was handled in the economic model via sensitivity analyses (Section 8).

Additional Recent Evidence—Diagnostic Accuracy

The results of the first multi-center evaluation of the diagnostic accuracy of CCTA, the Assessment by Coronary Computed Tomographic Angiography of Individuals UndeRgoing InvAsive Coronary AngiographY (ACCURACY) study, were very recently published (Budoff, 2008). In this study, data were obtained from 16 US sites, and included 230 patients who were referred for non-emergent ICA and also received CCTA. Certain exclusion criteria common to smaller validation studies (e.g., obesity, high calcium scores, vessel size) were not employed, as the study was designed to enroll a population similar to what might be expected in typical practice. The prevalence of CAD on ICA was 24.8%; mean (SD) age was 57 (10) years, and 59% of patients were male. Using a CAD threshold of ≥50% stenosis, patient-based sensitivity (95% [95% CI: 85%, 99%]) and specificity (83% [76%, 88%] reported in this study were very similar to our pooled estimates. The study also employed an alternative definition of ≥70% stenosis; diagnostic accuracy was essentially identical to that observed in primary analyses.

Another very recent publication reports findings from a second multi-center diagnostic accuracy study, the Coronary Artery Evaluation Using 64-Row Multidetector Computed Tomography Angiography (CORE 64) study (Miller, 2008). This study obtained data from a total of 9 sites in 7 countries, and included 291 patients aged 40 years and older (median: 59; 74% male) who were suspected of having CAD; patients first underwent CCTA and received ICA within 30 days. Unlike the ACCURACY study, potential CORE 64 participants were excluded if BMI exceeded 40, and if calcium score was 600 or higher; vessels less than 1.5 mm in diameter were also excluded. The prevalence of CAD on ICA was 56%. Findings from this study differed from both ACCURACY and our meta-analysis: patient-based visual CCTA sensitivity was 83% (95% CI: 76%, 88%), and specificity was 91% (85%, 96%). Potential reasons for this discrepancy (e.g., variability in technical competence, site dominance, spectrum bias) were not discussed in the study report.

A second iteration of the ICER meta-analysis was conducted with both of these multi-center studies included. There was no material change in our pooled estimates (sensitivity dropped from 97% to 95%, specificity was unchanged); however, in all likelihood because of the discrepant findings between these two large studies, statistical heterogeneity increased substantially. Because there are no available data on the possible reasons for these divergent results, we did not change our primary meta-analysis; results of analyses that included the multi-center studies are available in Appendix H.

Other Evidence

While our review focused on prospective studies, findings from several retrospective studies have recently appeared in print. Chang and colleagues compared clinical and financial records of 643 patients requiring testing to rule out ACS in the University of Pennsylvania Health System by (a) immediate CCTA alone; (b) observation unit plus CCTA and biomarkers; (c) observation unit plus stress testing and biomarkers; or (d) inpatient

admission with biomarkers and hospitalist-directed evaluation (Chang, 2008). Patients were matched by age, gender, race, TIMI score, and initial EKG. At day 30 of follow-up, patients in the CCTA alone group had no deaths, MIs, or readmissions, all of which were statistically significant in comparison to the other groups.

Two reports of retrospective analyses of healthcare claims also were identified. Both studies involved comparisons of patients without known CAD who underwent CCTA and were matched on selected patient characteristics to patients receiving SPECT. Findings from the first study, which evaluated clinical outcomes and costs for 1,938 and 7,752 CCTA and SPECT patients respectively over a 9-month period (Min, 2008 [1]), indicated no significant differences in the rate of CAD hospitalization, CAD outpatient visits, myocardial infarction, or new-onset angina between groups.

Results from the second study, which analyzed data on 1,647 CCTA and 6,588 SPECT patients over one year of follow-up (Min, 2008 [2]), indicated that CCTA was associated with a significantly lower rate of new-onset angina (4.3% vs. 6.4%, p<.001) and a reduced risk of angina or MI at one year (hazard ratio 0.70; 95% CI 0.55, 0.90); the rate of percutaneous or surgical intervention was also significantly reduced (0.2% vs. 0.8%, p<.001). No significant differences were observed in the rate of CAD hospitalization.

Studies with Relevant Subgroup Data

Stratified by CAD Risk or Pretest Likelihood

One common criticism of the existing diagnostic accuracy studies of CCTA is that the populations examined tend to be at higher risk for underlying CAD than will be patients that are likely to receive the test in practice (Budoff, 2006). Two studies in our sample address this issue by stratifying the population according to risk or pretest likelihood of CAD:

- Husmann et al. (2008) [1]: A total of 88 consecutive patients with suspected CAD were scheduled for both CCTA and ICA; patients were stratified into low, intermediate, and high risk categories based on Framingham risk score. In this population, which had an overall CAD prevalence of 49%, findings suggested that CCTA performance at ruling out disease was similar across risk categories (sensitivity 90.0%, 87.5%, and 100.0% for low, intermediate, and high risk respectively, p=.33; NPV 95.0%, 85.7%, and 100.0%, p=.45); a trend toward higher positive predictive value was observed, however, with increased levels of risk (PPV 64.3%, 93.3%, and 89.5% for low, intermediate, and high risk respectively, p=.07).

- Meijboom et al. (2007) [1]: In one of the largest studies reported to date, a total of 254 patients referred for ICA in the Netherlands received CCTA within one week prior to or following CCTA. Pretest likelihood of CAD (i.e., low, intermediate, or high) was estimated for each patient using the Duke Clinical Score. Overall prevalence of CAD on ICA was 50%. Sensitivity and NPV were similar across the three groups; consistent with findings from Husmann, there was a trend toward lower specificity (93%, 84%, and

74% for low, intermediate, and high) and higher PPV (75%, 80%, and 93%) as pretest likelihood increased.

Stratified by Gender
There has been considerable debate regarding the diagnostic performance of non-invasive CAD testing in men vs. women; some studies have suggested a greater challenge in women (Bairey Merz 2006), while others have found no differences (Gibbons 2002; Klocke 2003). Regardless, gender-based differences in anatomy, exercise tolerance, heart rate, level of coronary calcium, and other factors have led to continued interest in examining the influence of gender on diagnostic test results. Two studies have examined this issue with respect to CCTA:

- Pundziute et al. (2008): A total of 103 consecutive patients (51 male, 52 female) presenting with either known (34% of sample) or suspected CAD at Leiden University Medical Centre (Leiden, the Netherlands) were scheduled for ICA and received CCTA within a median of 4 weeks. Findings from this study suggested no material differences by gender in any measure of diagnostic accuracy.

- Meijboom (2007) [2]: In a larger sample from the same institution described above, a total of 402 patients (279 men, 123 women) scheduled for ICA (approximately 10% of whom had prior known CAD) received CCTA within one week. In this study, sensitivity and NPV were at or near 100% for both men and women; however, specificity (90% vs. 75%) and PPV (95% vs. 81%) were significantly greater in men.

Incidental Findings
A controversial feature of CCTA is its concurrent ability to detect abnormalities outside the heart; in particular, pulmonary nodules have been frequently reported as incidental findings of CCTA, likely due to both the adjacency of the pulmonary anatomy and the presence of standardized criteria for following "significant" nodules (MacMahon, 2005). Incidental lesions present a clinical and policy challenge because of the possible benefits of early detection of a small percentage of significant lesions relative to the costs and risks associated with further investigation of the majority of incidental findings whose identification and even treatment would be unlikely to provide a net health benefit to the patient.

We reviewed the current literature for studies that reported extra-coronary findings with multi-slice CCTA; because there are very few data from studies using 64-slice technology, we also reviewed studies based on earlier-generation multi-slice scanners. The results of our review are summarized in Table 5. Any summary of this literature is complicated by differing definitions of "clinically important" lesions, as these are typically based on the consensus of reviewing physicians. The reported rate of patients with *any* detected lesion ranged from 15% to 80%; "clinically important" lesions presumed to require follow-up have been found in 5-20% of patients evaluated. An unusually high percentage of clinically important findings (56.2%) was reported in a recent series of 258 Israeli patients (Gil, 2007); these results were primarily manifested in pulmonary nodules, however, and this study

featured both a lower cutoff for clinical significance of these nodules (>4 mm) and a higher percentage of current smokers than the other series analyzed. That said, current guidelines do suggest at least one further scan for even small nodules (MacMahon 2005). In addition to pulmonary nodules, the most common lesions deemed "clinically important" include thoracic or abdominal aortic aneurysms, pulmonary emboli, pleural effusion or infection, and hepatic or abdominal masses.

Despite the reported range and variability in defining clinical importance, it appears that relatively few lesions reveal significant pathology upon further investigation. In the largest series reported to date (Cademartiri, 2007 [4]), 81/670 (12.1%) patients had significant findings deemed to require follow-up or further investigation. Among these patients, 2 had newly-discovered pathologies (one pulmonary embolism and one bony metastasis from renal carcinoma). In another large series (Onuma, 2006), 114/503 (22.7%) had clinically-significant findings; upon subsequent review of medical records, a total of 18 patients (3.6%) were found to have therapeutic consequences (i.e., further treatment was required) from these incidental findings, and 4 patients (0.8%) had newly-discovered malignancies.

None of the studies we reviewed attempted to estimate the costs of further investigation of incidental findings on CCTA. We discuss the potential short-term economic impact of incidental findings in the economic model component of this report (see Section 8).

Although incidental findings are not an issue for stress EKG or stress echocardiogram, a recent case series involving 582 consecutive patients undergoing myocardial SPECT imaging with a Tc-99m sestamibi tracer (Gedik, 2007) reported extra-cardiac findings in 7 patients (1.2%). These were noted via either increased or decreased extra-cardiac uptake of the tracer, and included cases of thymoma, goiter, and sarcoidosis.

Harms

Other than small percentages of patients who did not complete the CCTA exam because of refused consent or psychological reactions (e.g., claustrophobic reaction), no studies reported immediate adverse events directly due to CCTA. This is likely because the most common expected event (reaction to contrast media) was mitigated by excluding patients with known allergies or reactions to contrast media as well as those with compromised renal status. In general, the incidence of severe or permanent reaction to contrast media is low.

While a recent examination of the use of prophylactic measures to reduce contract-induced renal injury (Weisbord, 2008) indicated that the incidence of elevated serum creatinine ranged from 0-11% after CT examination (depending on the threshold employed to indicate injury), this biochemical change was not independently or significantly associated with hospitalization or death. Findings from a meta-analysis of over 300,000 parenteral administrations of contrast media (Caro 1991) estimate the incidence of severe reactions or death at <0.01%. More recently, the renal effects of CCTA in 400 patients with chronic renal insufficiency was examined (El-Hajjar, 2008); the incidence of contrast-induced nephropathy was low (1.75%), and no patient required hemodialysis.

Radiation Exposure and Future Cancer Risk

Potential adverse health effects associated with radiation exposure are important factors to consider in the evaluation of CCTA as a potential diagnostic tool in the ED and/or outpatient settings, particularly because patients may already be exposed to radiation at other points along the diagnostic pathway (e.g., ICA, SPECT). Radiation dose is a measure of ionizing energy absorbed per unit of mass, expressed as units of Gy (Gray) or mGy; it often is quoted as an equivalent "effective" dose, in units of Sv (Sievert) or mSv. For x-rays, the radiation type produced by CT scanners, 1 mSv = 1 mGy. To place the effective radiation dose received from CCTA in context, the reported range of radiation in our sampled studies is listed in the table below along with typical doses from other tests and exposures to x-rays. Note that the doses received from ICA are similar to those at the lower end of the reported range for CCTA, while the range of SPECT doses are similar to those at the higher end of the reported range for CCTA:

Radiation exposure scenario	Approximate effective dose (mSv)
Chest x-ray	0.02
Round-trip flight, New York-Seattle	0.06
Low-dose CT colonography	0.5-2.5
Lumbar spine x-ray	1.3
Head CT	2.0
Single-screening mammogram (breast dose)	3.0
Annual background dose caused by natural radiation	3.0/yr
CCTA (lower reported range)	**2.0-8.0**
Invasive coronary angiography	***5.0-7.0***
Adult abdominal CT scan	10.0
Single photon emission CT (SPECT): Technetium	***9.0-13.0***
CCTA (higher reported range)	**12.0-14.0**
Typical dose to A-bomb survivor at 2.3 km distance from ground zero Hiroshima	13.0
SPECT: Thallium	***15.0-20.0***
Annual radiation worker annual exposure limit	20.0/yr
Annual exposure on international space station	170.0/yr

Sources: Brenner, 2005; FDA [www.fda.gov/cdrh/ct/risks.html]; ICER CCTA systematic review; Van Gelder 2004, Mettler 2008, Shuman 2008; Earls 2008; Husmann 2008 [2].

The primary risk associated with exposure to ionizing radiation is cancer. According to the FDA, estimates based on the experience of A-bomb survivors suggests that a dose of 10 mSv may be associated with an increase in the possibility of fatal cancer of approximately 1 chance in 2000. This risk level is relatively small in comparison to the approximately 400 out of 2,000 individuals expected to develop cancer from all other causes combined. Dose levels for all of the above-listed diagnostic tests are "moving targets"; attempts to reduce radiation dose are not specific to CCTA.

There is considerable controversy on extrapolating cancer death risks from those experienced by adults with high radiation exposure at Hiroshima to the potential risks at much lower radiation doses. However, linear extrapolation has been the approach generally used, although the uncertainties inherent in this approach become progressively greater at lower doses. Also controversial is whether a natural threshold of radiation exposure exists before excess risk from specific exposures can be realized. The current guidance from a variety of regulatory authorities is that no threshold exists, but this has also been intensely debated.

Our evidence review found 17 articles in which the radiation dosage was estimated. Estimated radiation dosages for CCTA ranged widely, from 4.6 to 21.4 mSv. In general, the lowest rates in the reported range were from studies employing dose-sparing protocols such as tube current modulation (see discussion below) (Ropers D, 2006, Nikolaou, 2006) as well as those using dual-source scanners (Johnson [2], 2007, Leber, 2007).

In general, calculated radiation doses were higher in women (range: 10.24-21.4 mSv vs. 7.45-15.2 mSv in men), due to the higher density of breast tissue in women. These estimates do not differ materially from those reported elsewhere in the literature for CCTA, which range from 5-32 mSv and average 16 mSv (Mettler, 2008).

Most of the studies we reviewed employed some form of dose-sparing protocol to attempt to reduce radiation exposure. The most common of these was prospective EKG gating, in which the heart is only scanned at certain times during the cardiac cycle, so the patient does not receive radiation during the entire examination (Healthcare Human Factors Group, 2006). In some settings, prospective EKG gating has been found to reduce average effective doses to 2-4 mSv (Shuman, 2008, Earls, 2008, Husmann, 2008 [2]); however, results from a recent presentation of data from a multicenter study suggest that effective doses still vary widely (reported range: 5-37 mSv) by institution, even with over 80% of centers employing prospective gating protocols (Hausleiter, 2008).

Other techniques to reduce radiation exposure from CCTA include automatic exposure control, in which the tube current is adjusted to the anatomy of the patient, and the so-called "step-and-shoot" strategy, in which images are acquired at predetermined stop points during the scanner's spiral revolution. In addition, it is thought that the introduction of 256- and 320-slice scanners may further reduce exam time; whether this leads to a net reduction in radiation dose is unclear, as the higher precision of the newer machines may deliver increased radiation at the outset.

In an attempt to examine the attributable radiation-induced cancer risk from CCTA, a recent analysis used Monte Carlo simulation methods applied to mathematical phantom data on organ doses to men and women during 64-slice CCTA (Einstein, 2007). Findings indicated that the lifetime attributable risk for cancer was low but non-negligible (0.22% and 0.08% in women and men aged 60 years respectively); prospective EKG gating would be expected to reduce this risk by about 35%.

Prior Published Studies on Economic Impact of 64-Slice CCTA
Limited data are available on the potential economic impact of CCTA in coronary artery disease; studies that have been published are based on decision-analytic models or retrospective database analyses. The studies vary widely in terms of their structure, strategies evaluated, and assumptions about test characteristics and costs. Accordingly, direct comparison of the findings is difficult.

ED Studies
The model that formed the general structural basis for our ED model (see Section 8) has been previously published (Ladapo, 2008); in this model, patients presenting with chest pain (underlying prevalence of cardiac chest pain=12%; prevalence of CAD=27%) were evaluated alternatively with CCTA in addition to standard ED triage care (i.e., serial enzymes, stress testing, observation) or standard care alone. Separate strategies for men and women were evaluated; costs were estimated on a lifetime basis, and utilities for long-term outcomes of appropriate and inappropriate diagnosis were incorporated. Findings suggest that CCTA would be cost-saving in women and would generate slightly increased costs in men. On a lifetime basis, CCTA would dominate standard care in women and have an incremental cost-effectiveness ratio (ICER) of $6,400 per QALY in men.

Another recently published study (Khare, 2008) also examined CCTA's cost-effectiveness on a lifetime basis, in a population with very low CAD prevalence (6%). The competing strategies in this analysis (CCTA or standard care) produce either positive, indeterminate, or negative results; no distinction is made between "significant" or "mild" stenosis on CCTA, and all positive results result in referral to ICA. Results indicated that CCTA was cost-saving relative to standard care, regardless of whether stress ECHO or stress EKG was the modality used for functional testing.

Finally, results of the recent retrospective evaluation of patients triaged in the University of Pennsylvania Health System described previously (Chang, 2008) suggested that immediate CCTA alone was associated with reductions in observation time of 12-22 hours and cost reductions of $1,100-$2,800 when compared to the other triage strategies employed.

Outpatient Studies
Two decision-analytic models have examined CCTA's cost-effectiveness in an outpatient setting (Mowatt, 2008, Dewey, 2007). Mowatt and colleagues used the structure of a previous model examining the cost-effectiveness of SPECT (Mowatt, 2004) to estimate the effects and costs of multiple single- and dual-test strategies during both the diagnostic phase and over a lifetime horizon. Other strategies involved stress EKG and stress SPECT; in addition, two strategies examined the impact of having CCTA be the final test in the diagnostic pathway (rather than ICA). All positive or indeterminate findings in these strategies result in a subsequent test or confirmation, and all negative results stop the testing flow. At a CAD prevalence level identical to our model (30%), the most effective strategies are CCTA-ICA and ICA alone, while the lowest-cost strategies are stress EKG-CCTA and CCTA alone. In lifetime modeling, comparison of the strategies involving CCTA indicated that a stress EKG-CCTA-ICA strategy is a cost-effective alternative relative

to the stress EKG-ICA (£9,200 per QALY gained) and stress EKG-CCTA (£1,400 per QALY gained) strategies. In addition, all CCTA strategies were dominant in comparison to strategies involving SPECT.

In the other model (Dewey, 2007), a total of 6 strategies were evaluated for patients presenting with stable chest pain: CCTA, calcium scoring using electron-beam CT, dobutamine stress MRI, stress EKG, stress ECHO, or immediate ICA. Multiple hypothetical cohorts were evaluated according to different pretest likelihoods of disease. As with the Khare model described above, this analysis assumes that *all* positive findings on the initial test are referred for ICA. Cost-effectiveness was expressed in terms of cost per correctly identified CAD patient; this appears to have been generated as a "stand-alone" result for each strategy, however, and was not evaluated incrementally among the strategies. CCTA generated the lowest cost per correctly identified patient at pretest likelihoods of 10-50%; ICA (which was assumed to be 100% accurate) performed best at pretest likelihoods of 70% or higher.

Costs were also analyzed in the two retrospective claims studies previously described (Min, 2008 [1,2]). In the first, 9-month CAD-related costs were $445 lower for the CCTA group vs. SPECT after multivariate adjustment for demographic characteristics, pre-test expenditures, comorbidities, and cardiac medication use. In the second, unadjusted CAD-related expenditures were significantly lower for CCTA at one month and 6 months; differences remained at one year ($3,542 vs. $4,605, $p<.0001$), even after multivariate adjustment.

7.4 Summary

The body of prospective published evidence on the impact on patient outcomes of CCTA as part of a diagnostic strategy compared to usual care is limited to 7 case series and a single RCT, all but one of which were evaluated in the ED setting. The results of one study (Rubinshtein, 2007 [3]) demonstrated a significant impact on clinician decision-making in the ED through which CCTA reduced hospitalization and additional procedures in many patients while having no adverse outcomes among patients discharged home. These findings have not been confirmed by other studies or explicit comparisons to other diagnostic strategies, save for two recent retrospective claims-based studies. While these studies suggest that cost savings and some clinical benefit are achievable for CCTA vs. SPECT, attendant selection and other biases common to such quasi-experimental research place an important qualification on their results.

The literature on the diagnostic accuracy of CCTA vs. ICA has expanded rapidly over the last three years, and with notable consistency the evidence suggests that CCTA has a very high sensitivity (~97%) for significant occlusion and a moderately high specificity (~82% if non-evaluable scans are considered false positives, ~87% if such scans are excluded from consideration). These data have been generated in patient populations around the globe, often among patients with relatively high underlying prevalence of CAD, raising questions about the applicability of findings to patient populations at low-to-intermediate (10-30%) risk of CAD. Studies of diagnostic test accuracy can suffice if clinicians already have

evidence from randomized trials showing that treatment of the cases detected by the diagnostic test improved patient outcomes, but the body of evidence on CCTA does not yet include studies to address this question.

There are a number of other questions that the current evidence does not address. For one, the lack of data on long-term outcomes with CCTA makes it difficult to ascribe value to its ability to reduce the rate of false-positive and false-negative findings relative to other strategies. Without these data, we do not know whether and when false negatives will re-present with symptoms and be diagnosed correctly, and whether they will suffer any health consequences in the intervening period. It is also impossible to know whether medical treatment of false positives would provide a net health benefit given that CAD will develop over time in many healthy individuals.

What is also unknown is whether the widespread adoption of CCTA will result in a shift in the distribution of candidates for such a strategy – for example, use of the test in very low risk individuals may shift the balance of true vs. false positives, thereby raising uncertainty as to its benefits on a population-wide basis; this uncertainty is particularly heightened in light of the unanswered questions around risks associated with CCTA's radiation dose as well as the health-system impacts of extra-coronary findings.

Also, because of CCTA's visual precision, "mild" levels of stenosis (i.e., 20-70%) can be detected; the benefits of aggressive management of this level of CAD are unknown, however, as such levels of stenosis cannot be directly linked to coronary insufficiency. While not a focus of our systematic review, several studies have attempted to examine CCTA's ability to diagnose functional cardiac deficits, using SPECT or another functional test as a reference (Gaemperli, 2007, Gallagher, 2007, Schuijf, 2006 [2]). While negative predictive value for these abnormalities was similar to that reported in the ICA-reference studies, positive predictive value ranged between 50-60%. Some have posited that, with increasingly precise technology, the ability to use CCTA to study blood flow and perfusion deficits will be heightened; evidence has not yet accumulated to support this, however.

Others argue that one of CCTA's utilities is in identifying so-called "vulnerable plaque"— i.e., coronary plaque that is at highest risk for rupture and formation of thrombi that cause acute cardiac events (Ambrose, 2008). Because CCTA's technology can be used to quantify the amount of calcified plaque (i.e., the "calcium score"), which has been cited as one of the risk factors in determining vulnerable plaque, some feel that detection of CAD in this earlier state would lead to more informed and efficient treatment decisions, reducing downstream risks and costs to the patient. The concept of vulnerable plaque is controversial in and of itself, however, as there are no current data on its natural history-- the rate of plaque progression, the characteristics associated with rupture, and the association with the incidence and timing of cardiac events are therefore unknown (Lau, 2004). Until such data are made widely available, the utility of CCTA in preventing the progression of early CAD will be speculative.

CCTA is a safe procedure; the immediate risks of the procedure itself are similar to those of other tests employing contrast media. The potential for harm from radiation, while modulated to some extent by the use of dose-sparing protocols, is still felt by some experts and commentators to be a significant concern , particularly if CCTA is being considered for use in combination with other radiation-based diagnostic tests (Einstein, 2007). However, there are many unanswered questions about the true risk function from test-induced radiation, and the role of radiation exposure in determining the net health benefits from CCTA will rely largely upon decision-maker values and judgment.

With CCTA the patient has the benefit of, but also potential harm from, extra-coronary findings. Clinically significant findings found during CCTA provide for early detection of a serious condition in some patients. Whether early detection leads to more effective treatment and improved outcomes cannot be determined from the available evidence. Similarly, there are no studies of the unnecessary expenses, inconvenience, and health risks attendant upon follow-up of less serious incidental findings.

Several large clinical studies are underway that may address concerns regarding CCTA's impact in clinical practice. Four RCTs are ongoing, all of which include major cardiovascular events as primary endpoints. In addition, a within-subject study sponsored by GE Healthcare is evaluating CCTA's diagnostic performance relative to SPECT as well as evaluating its impact on major cardiovascular events and the rate of downstream cardiac testing. Finally, a large observational study is underway at Brigham & Women's Hospital, Boston, following patients who are referred for stress perfusion with SPECT or PET, CCTA, or combined perfusion/anatomy studies; the primary endpoint of interest is referral for cardiac catheterization, as well as major cardiac events and the relative cost-effectiveness of each approach.

8. Decision Analytic/Economic Models

Objectives
The objectives of this decision analysis were to evaluate the clinical and cost-effectiveness of coronary computed tomographic angiography (CCTA) for the detection of coronary artery disease (CAD). Following the guidance of the ICER Evidence Review Group, the modeling was targeted to evaluate the use of CCTA for the following applications:

1. CCTA in the emergency department (ED) triage for patients with acute chest pain of unknown origin and a low-to-intermediate risk of acute myocardial infarction or unstable angina

2. CCTA as an outpatient screening tool for CAD in a low-to-intermediate risk population presenting with stable chest pain

Overview of Models
Because the clinical scenarios and patient populations related to CCTA use differ substantially between the ED and the outpatient settings, we decided to build two separate models that would most appropriately reflect the current standard of care and evaluate options for how CCTA could be introduced into these two settings.

The model evaluating CCTA for patients with acute chest pain in the ED setting loosely follows the algorithm of the RCT by Goldstein (Goldstein, 2007) such that in the CCTA branch, the detected luminal diameter of the stenosis determines further action for revascularization independently of the number of affected vessels (Ladapo, 2008).

The model evaluating CCTA as a tool for evaluating stable chest pain in the outpatient setting follows the CAD treatment recommendation derived from the recent COURAGE trial (Boden, 2007) and thus requires that the diagnostic tests not only identify stenoses correctly but also differentiate between 3-vessel/left main artery disease and 1- or 2-vessel disease. Both models will be described in more detail in the following sections.

In neither model are the potential benefits, harms, or costs of incidental findings included. This decision was made due to the lack of data describing the downstream balance of benefits and harms accrued through the identification and treatment of incidental findings. In addition, there is no consensus among clinical and policy experts on the likely balance of benefits and harms. Nonetheless, we did attempt to estimate the incidence of pulmonary nodules >4 mm in size, based on age- and gender-based data from the National Cancer Institute's Cancer Intervention and Surveillance Modeling Network (CISNET) Lung Policy Model (http://www.cisnet.cancer.gov/profiles) and the follow-up recommendations of the Fleischner Society (MacMahon, 2005). Briefly, the incidence of such nodules was estimated to be 19.8%, which we reduced by 30% (13.9%) due to the fact that CCTA visualizes approximately 70% of lung volume (Kirsch, 2007). We estimated follow-up costs based on Medicare reimbursements for the tests depicted in the guidelines, and arrived at a blended

average rate of approximately $700 for follow-up of nodules 4-8 mm and >8 mm in size. We applied this cost estimate in a post hoc analysis, examining their impact on costs in the ED setting.

Our decision analytic models also do not explicitly attempt to model long-term consequences of radiation exposure. This decision was also determined by the lack of data with which to estimate the incidence and distribution of possible radiation-induced cancers attributable to CCTA. In the outpatient model we report the number of patients who would be exposed to *any* radiation during the diagnostic testing.

We adopted a payer perspective for our evaluation; as such, cost estimates were largely based on CPT codes and national Medicare reimbursement as well as other studies. All costs were converted to 2008 US dollars using the medical care component of the Consumer Price Index. Following the current recommendation of the US Panel on Cost-Effectiveness in Health and Medicine, both costs and health outcomes were discounted at 3% annually (Gold, 1996).

8.1 ED Model

Overview

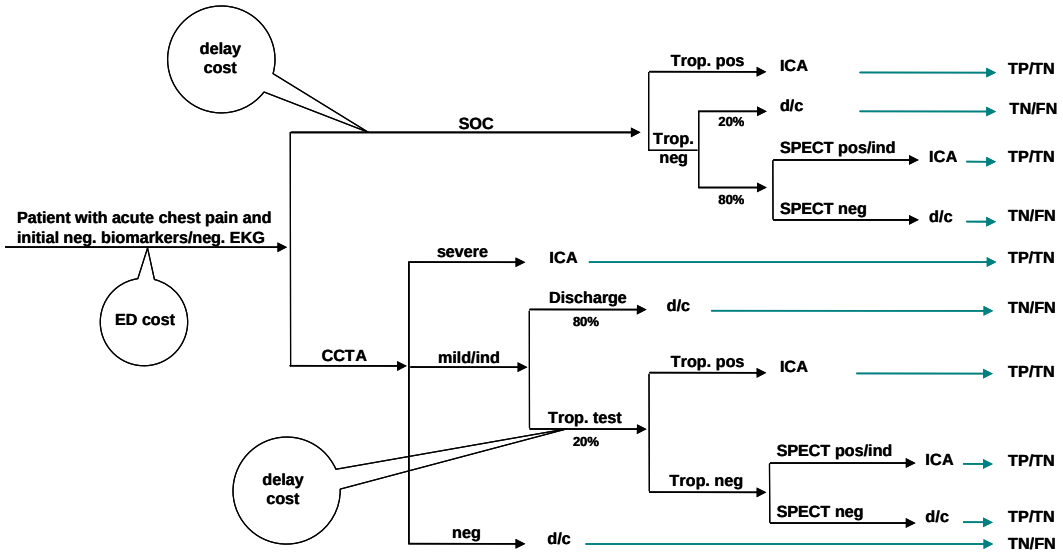
We modified a recently published microsimulation model, developed by Ladapo (Ladapo, 2008), to compare the diagnostic results of standard of care (SOC) to CCTA-based management in the triage of 55 year-old men with acute chest pain and at low-risk of an acute myocardial infarction or unstable angina. The model begins with a cohort of patients presenting to the ED with acute chest pain of unknown origin, initial negative biomarkers, and non-significant EKG changes.

Figure 1 on the following page depicts the possible pathways of the two strategies: In the SOC pathway, patients are re-evaluated with serial enzymes after 6-8 hours, and incur observation unit costs (i.e., "delay") while awaiting test results. Patients with elevated follow-up biomarkers are directly referred for invasive coronary angiography (ICA); among those with negative biomarkers, we solicited expert opinion on the percentage of patients who would then be discharged vs. the percentage referred for stress testing, and arrived at an estimate of 20% for discharge and 80% for stress testing. SPECT was selected for the SOC pathway upon the guidance of clinical experts; however, alternative analyses were conducted with stress ECHO as the standard test. Patients who have a SPECT that suggests a severe stenosis (≥50% for left main or ≥70% for vessels) or those with indeterminate test results are referred for ICA; patients with negative SPECT are discharged without further testing or treatment.

In the CCTA pathway, CCTA is integrated into the standard of care triage: during the waiting period for the follow-up enzymes, patients are imaged and either discharged, evaluated with a stress test, or sent directly to ICA depending on the severity of their atherosclerosis as suggested by CCTA. If CCTA reveals severe stenosis (≥50% for left main or ≥70% for any other vessel or vessels), the patient is immediately referred to ICA; if CCTA

reveals no stenosis, the patient is immediately discharged. If CCTA reveals a mild stenosis (<50% for left main or <70% for any other vessel or vessels) or the result is indeterminate, we assumed that 80% would be found to have non-significant CAD and be discharged; the remaining 20% would enter the standard triage pathway. While this distribution is an estimate without empirical foundation, it attempts to model the likelihood that patients with 20-30% occlusion or less would be considered for immediate ED discharge with pharmacologic treatment alone. We also conducted alternative analyses in which all patients with mild or indeterminate results would receive standard triage care.

Figure 1. ED Model Pathways



Notes: severe stenosis: 50% to 100% decrease in luminal diameter; mild stenosis: 1% to 49% decrease in luminal diameter; SOC: standard of care; CCTA: CCTA: coronary computed tomographic angiography; Trop.: troponin; ICA: invasive angiography; SPECT: Stress-single photon emission computed tomography

Because ICA is considered to be a gold standard, it will reveal the patient's true disease status. As a result, patients who undergo ICA will always be correctly diagnosed as having a severe stenosis that requires invasive treatment (true positive) or not having a severe stenosis (true negative). Patients discharged without receiving ICA can either be correctly (true negative) or incorrectly (false negative) diagnosed as free of any severe stenosis.

Input Parameters

Clinical Parameters

To evaluate the effectiveness of CCTA as a diagnostic instrument for the work-up of acute chest pain patients, two distributions amongst this population are essential parameters: the distribution of acute coronary syndrome (ACS) and non-ACS diagnoses and the distribution of coronary atherosclerosis within these diagnostic categories. All data were derived from the published literature and parameters were estimated as described by

Ladapo (Ladapo, 2008) and explained further in the following paragraph. All parameters are provided in Table I.

The distribution of ACS and non-ACS diagnoses in the initial ED visit (Table I) was derived from several studies that totaled more than 1,000 acute chest pain patients who had no history of heart disease and were considered to be at low risk for ACS based on a clinical algorithm constructed by Goldman and colleagues (Goldman, 1988; Zalenski, 1997; Sallach, 2004). Patients were assumed not to suffer from life-threatening conditions other than ACS. Although such patients may be experiencing aortic dissections, pulmonary embolisms, and other serious conditions, our omission of these health events likely does not impact incremental cost-effectiveness, as they would be evaluated similarly under both strategies.

The distribution of coronary atherosclerosis within the ACS and non-ACS diagnoses were derived from a large cohort of patients with chest pain who underwent invasive angiography but were not diagnosed with ACS (Chaitman, 1981). This source was selected because it came from a very large national database (the CASS study) and provided data on the underlying distribution of atherosclerosis within diagnostic categories similar to those used to characterize chest pain in the ED. Patients were stratified by age, gender, and their type of chest pain complaints being "definite angina," "probable angina," or "non-specific chest pain".

Using the Chaitman prevalence data, patients in our model with ACS were assigned a distribution of vessel disease similar to the "definite angina" chest pain group; patients with stable angina were assigned a distribution of vessel disease that averaged results from the "definite angina" and "probable angina" groups, as we assumed these patients were healthier than patients with ACS; patients with non-cardiac chest pain were assigned a distribution of vessel disease similar to the "non-specific chest pain" group.

As shown in Table I on the following page, the majority (88%) of all acute chest pain patients in the ED experience non-cardiac related chest pain. However, the ICA data from Chaitman demonstrated that among 55-year-old men there is a total prevalence of severe stenoses of 27% and a prevalence of mild stenoses of 28%. Thus our model assumes that some patients will present to the ED with non-specific chest pain due to other causes but who, if sent for stress echocardiography or CCTA, will ultimately be found to have at least one vessel with a stenosis ≥50%. This approach is the best way to create parameters for a model that reflects the clinical reality that results of CCTA lead to multiple pathways of further evaluation/treatment.

Table I: Patient and diagnostic test characteristics

Variable	Base Case Estimate	Source(s)
Initial distribution of disease in ED		
Non-ST segment elevation MI	0.03	Ladapo, 2008; Sallach, 2004; Zalenski, 1997
Unstable angina	0.07	"
Stable angina	0.02	"
Non-cardiac chest pain	0.88	"
64-slice CCTA characteristics		
Probability of classifying severe coronary stenosis as		
Severe	0.92	Shabestari, 2007; Zalenski, 1997
Mild	0.07	"
Normal	0.01	"
Probability of classifying mild coronary stenosis as		
Severe	0.21	"
Mild	0.72	"
Normal	0.07	"
Probability of classifying normal coronary arteries as		
Severe	0	"
Mild	0.02	"
Normal	0.98	"
Indeterminacy rate	0.03	ICER Review
Stress-SPECT		
Sensitivity for CAD	0.88	Garber, 1999
Specificity for non-CAD	0.77	"
Indeterminacy rate	0.09	Patterson, 1995
Serial troponin measurement		
Sensitivity for NSTEMI	0.95	Lau, 2001
Specificity for patients not having NSTEMI	0.90	Lau, 2001
Mortality from ICA	0.001	Kuntz, 1999

Notes: CAD = coronary artery disease, ED = emergency department, MI = myocardial infarction. NSTEMI: non-segment elevation myocardial infarction

This approach creates a model which, in comparison to the clinical experience of many physicians, will result in a very high proportion of ED chest pain patients with a positive

troponin test or SPECT who will subsequently be sent for ICA. This feature arises because the cohort of 1,000 patients includes those 10% who have unstable angina or who will develop MI; they also include 880 patients who present with "non-cardiac chest pain," but who, given that the cohort represents 55-year old men, have an underlying 18% prevalence of significant CAD. In addition, our model sends many patients with indeterminate SPECT or CCTA tests to further testing and/or ICA. When these features are combined it is not surprising to see relatively high total numbers of patients sent for ICA.

Test Accuracy

No published ED studies have reported all 64-slice CT coronary angiography test characteristics on a per-patient basis as required for the model, so Ladapo used data that applied to individual segments of the coronary arteries (see Table I). Note that this method of reporting will, on average, underestimate the diagnostic power of CCTA because many patients have multiple significant coronary lesions.

The diagnostic performance of other tests, including serial troponin measurements and SPECT for identifying coronary artery disease were derived from a published meta-analysis (Garber, 1999). Based on findings from ICER's systematic review, CCTA was assumed to provide non-diagnostic results at a rate of 3.2%, and patients with non-diagnostic exams were subsequently evaluated using the standard triage care paradigm.

Costs

ED costs were estimated using Medicare reimbursement data (Centers for Medicare & Medicaid Services, 2008). Table II on the following page depicts the detailed CPT codes associated with each cost item. To account for the costs of admission of patients to an ED observation unit when prolonged evaluation was required, we assumed that these "delay costs" would apply for all patients in the SOC strategy and for those in the CCTA strategy whose CCTA result indicates a "mild" stenosis and requires the patient to spend additional time undergoing further evaluation in the ED observation unit.

Table II: Cost Parameters

Procedure, CPT code (description)		Total costs ($)	Source
Delay	APC 0339	443	CMS, 2008
SPECT	78465 (heart image (3d), multiple) 78478 (heart wall motion add-on) 78480 (heart function add-on) 93015 (cardiovascular stress test)	765	"
CCTA	0145T (CT heart w/wo dye funct: $306) Physician fee ($159)	466	"
ICA	93508 (cath placement, angiography) 93510 (left heart catheterization) 93543 (injection for heart x-rays) 93545 (injection for coronary x-rays) 93555 (imaging, cardiac cath)	2,750	"
ED visit	Micro-costing study excluding costs for delay and diagnostic testing	890	Goldstein, 2007

Notes: Delay: delay cost attributed to those patients who are closely monitored for 6-8 hrs. as part of their diagnostic workup; SPECT: single-photon emission computed tomography; CCTA: coronary computed tomographic angiography; ICA: invasive coronary angiography; CMS: Centers for Medicare and Medicaid Services

Model Analyses

We ran a first-order Monte Carlo micro-simulation model and reported the average results for 1,000 patients. This model only considers the diagnostic results and reports the number of correctly diagnosed diseased patients with a severe stenosis requiring invasive intervention (true positives), correctly diagnosed patients without a severe stenosis (true negatives), and incorrectly diagnosed diseased patients (false negatives). Furthermore, the model reports the total number of ICAs performed the number of negative ICAs, and number of ICA-related deaths as well as the associated costs for both strategies. We also report the number of patients with incidental findings in the CCTA strategy who require diagnostic follow-up.

Results

Base Case Analysis

Table III on the following page depicts the results for a cohort of 1,000 55-year old men. The left hand column shows the result if all patients had undergone the SOC strategy and the right hand column depicts the results if the identical 1,000 patients had all undergone the CCTA strategy. Among the notable differences between SOC and CCTA + SOC are the numbers of false negatives (63 vs. 16), the number of cases of "missed" acute coronary

syndrome (18 vs. 5), number referred for ICA (434 vs. 327), and patients sent for ICA who return with normal coronary arteries (228 vs. 74). In addition, the CCTA + SOC strategy allows for the immediate discharge of 567 patients (vs. 0 in the SOC strategy; data not shown). ED testing costs are higher for CCTA + SOC, but when the savings of fewer angiographies and lower delay costs are factored in, an average savings of $719 per patient is observed for the CCTA + SOC pathway. When the costs of following the 14% of patients in the CCTA + SOC with incidental findings were included (about $100 per patient undergoing CCTA), cost savings were reduced to $619, but remained in favor of CCTA.

Note that the number of patients referred to ICA is higher than many clinicians would expect. The reason for this is twofold: the rather high underlying CAD prevalence of 27% results in 206 necessary ICAs for SOC and 253 for CCTA + SOC. In addition, the model includes two different paths leading to unnecessary ICAs: (1) false-positive test results for severe stenosis and (2) indeterminate test results, most of which are sent to ICA. Finally, while not depicted below, the CCTA + SOC strategy will expose all patients to radiation, vs. 43% in the SOC strategy.

Table III: Base case results

Outcomes (per 1,000)	SOC	CCTA + SOC
True positive	206	253
True negative	731	731
False negative	63	16
False negative results with ACS	18	5
Referred for ICA	434	327
ICA negative results	228	74
ICA related deaths	0.04	0.03
Incidental findings	0	138
Costs ($ per patient)		
ED/patient	1,421	1,409
Delay/patient	443	33
Cath lab/patient	1,193	898
Total/patient	**3,060**	**2,341**
Cost difference (CCTA vs. SOC)	**- $719**	

Notes: SOC: standard of care; CCTA: CCTA: coronary computed tomographic angiography; ACS: acute coronary syndrome

Alternative Analyses

When the test modality of SPECT was replaced with stress ECHO, there was a tradeoff of reduced sensitivity (76% vs. 88% for mild stenosis, 94% vs. 98% for severe stenosis) for better specificity (88% vs. 77%). As such, the number of true positives decreased, but this

was outweighed by a substantial decrease in the number of patients referred for ICA as well as the number with ICA-negative findings. Because these changes affected all patients in the SOC pathway but only a fraction of patients in the CCTA+SOC pathway, and due to higher testing costs with SPECT ($765 vs. $300 for stress ECHO), cost savings in the CCTA+SOC pathway were reduced to $314 under this scenario (see Appendix I for full results).

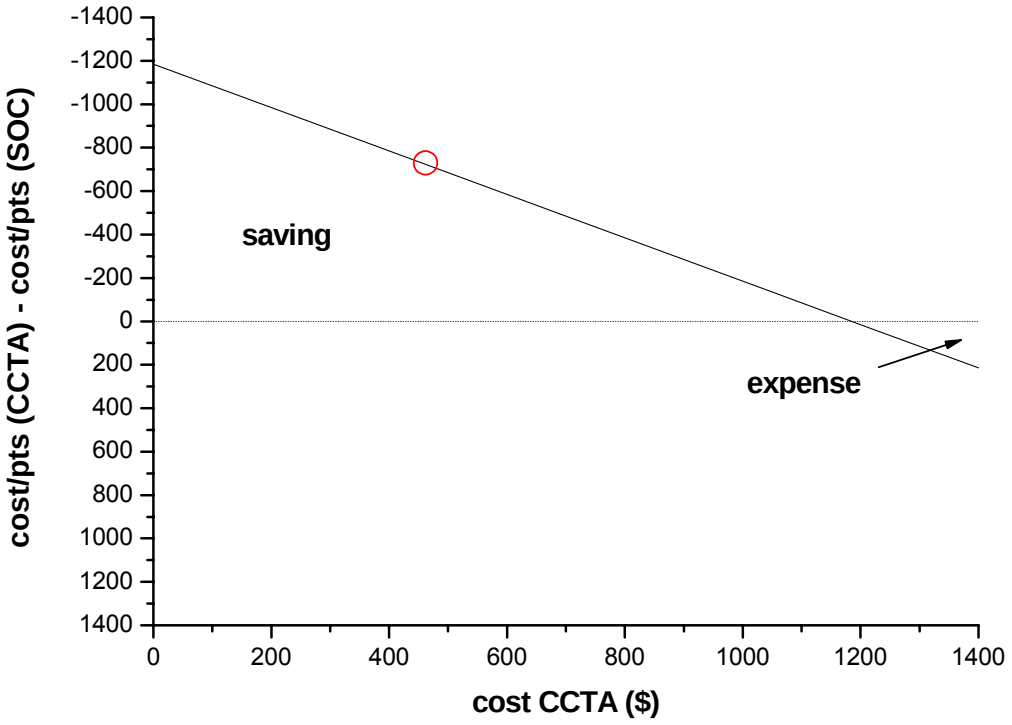
Sensitivity Analyses

Deterministic sensitivity analyses explore the effect that a change in one or more parameters over a plausible range of values will have on the results, in the case that all other parameters are held constant. This type of analysis is meant to answer 'what if' questions. We present the results of deterministic sensitivity analyses for the cost of CCTA and delay costs.

Sensitivity Analysis – Cost of CCTA

Costs of CCTA occur in the CCTA as one-time cost for all patients in this strategy and for no patients in the SOC strategy. For the base case, we assumed a cost of $466 resulting in an average cost-saving of $719 per patient. Figure 2 below depicts the linear relationship between CCTA costs and the cost difference between the two strategies. When CCTA costs $1,185 or less, CCTA is cost-saving compared to SOC.

Figure 2: Sensitivity analysis on costs of CCTA



Note: circle: base case estimate for CCTA cost.

Sensitivity Analysis – Costs of ED Delay
Delay costs occur as a one-time cost in both strategies for all patients who have to be carefully observed until they have received their serial enzyme and stress tests to rule in/out myocardial damage. These costs apply to all patients in the SOC strategy and to 20% of those patients in the CCTA+SOC strategy who have CCTA findings indicating a "mild stenosis" or indeterminate results. For the base case we assumed a cost of $443 resulting in an average cost savings of $719 per patient. When we attempted to conduct threshold analyses, there was no delay cost at which CCTA failed to be cost saving, even if these costs were set to zero; CCTA-associated savings in the model were therefore driven primarily by reduced need (and associated costs) for invasive angiography.

Conclusions
Our model therefore is consistent with other published cost-effectiveness analyses in suggesting that when used as part of a triage strategy for low-to-intermediate risk chest pain patients in the ED, CCTA will allow the more rapid discharge of nearly half of all patients and decrease the number of false negative diagnoses while reducing the number of angiographies compared to the current standard of care. According to the model CCTA is also cost-saving, with about $719 in savings per patient in comparison to SOC. Taking into account the additional follow-up costs for the 14% of patients who undergo CCTA and have incidental findings (approximately $100 per patient receiving CCTA), the cost-savings are reduced to approximately $619, but remain in favor of CCTA. However, CCTA does expose every patient to radiation, whereas only about 43% of the patients in SOC are exposed via invasive angiography.

8.2 Outpatient Model

Overview
We modified an existing microsimulation model that was initially developed by Joseph Ladapo MD, PhD, as part of his doctoral dissertation at the Harvard School of Public Health to assess CCTA in the evaluation of patients with stable chest pain, using conventional diagnostic modalities as comparators.

The base case population consisted of 55 year-old men with stable chest pain and with either low (10%) or intermediate (30%) risk of underlying significant CAD -- one or more vessels with occlusion ≥70% or left main occlusion at ≥50%. The model reported multiple outcomes for each strategy: the intermediate diagnostic results, expressed as numbers of correctly and incorrectly indentified patients with CAD, the number of resulting invasive angiographies, the number of patients exposed to radiation, the cost for diagnostic work-up, and the long-term prediction of remaining quality-adjusted life years and lifetime medical costs.

Diagnostic Phase

Diagnostic Strategies
We considered 8 different strategies, alone and in combination, in order to capture a wide range of management approaches for evaluating patients with stable chest pain and a low-to-intermediate risk of CAD:

1. Coronary Computed Tomographic Angiography (CCTA)
2. Stress-Echocardiography (Stress-ECHO)
3. Stress- Single Photon Emission Computed Tomography (Stress-SPECT)
4. CCTA followed by Stress-ECHO
5. Stress-ECHO followed by CCTA
6. CCTA followed by Stress-SPECT
7. Stress-SPECT followed by CCTA
8. Stress-ECHO followed by Stress-SPECT

Diagnostic Pathways
The model begins in an outpatient setting with evaluation of patients with stable chest pain and it is designed to differentiate between the management of three different test results reflecting different levels of CAD severity:

1) Negative for CAD
2) Positive for CAD (if a functional test) or Positive for one- or two-vessel CAD (if CCTA); and
3) Markedly positive for CAD (if a functional test) or Positive for 3-vessel or left-main artery disease (if CCTA)

Generally, the alternative diagnostic pathways differ between 1-test and 2-test strategies. In the 1-test strategy (Figure 3a on the following page), a single test is performed and patients with markedly positive test results or whose test results are indeterminate are sent for ICA. Depending on the ICA findings, patients can either be true positive or true negative for three-vessel disease or left-main disease (3VD/LM). True positives are treated with aggressive medical therapy and revascularized with coronary artery bypass (CABG) surgery.

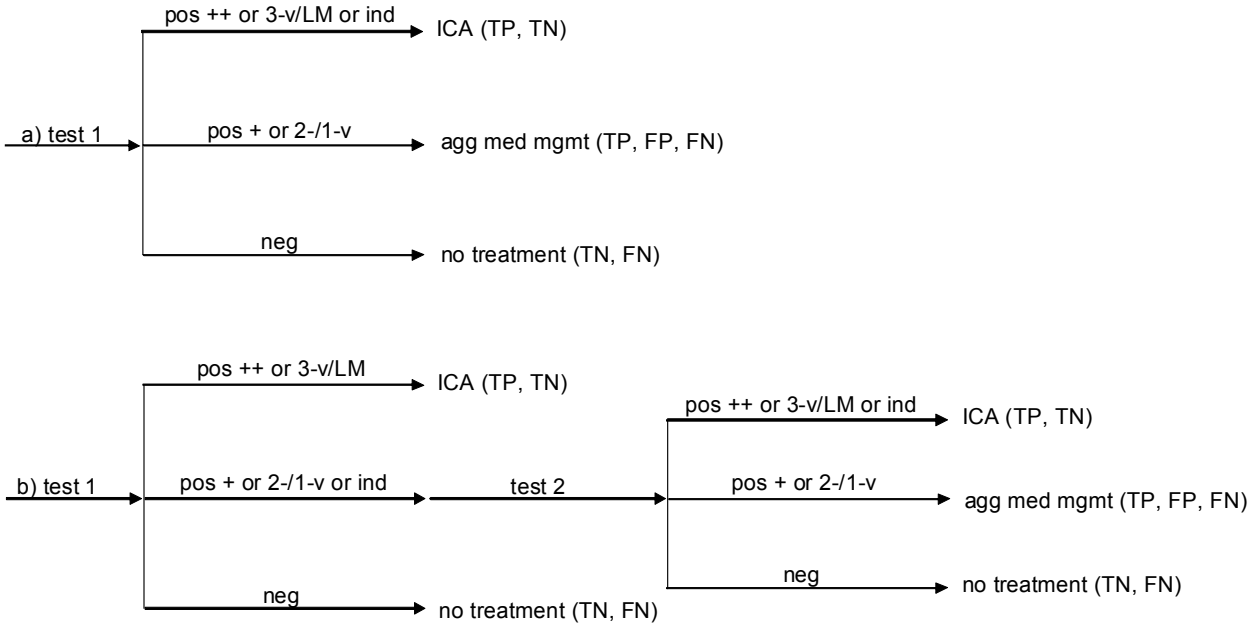
Patients whose diagnostic test is positive, but not markedly positive, for CAD are all started on aggressive medical treatment as per the treatment guidelines suggested by the COURAGE trial (Boden, 2007). As all non-invasive tests are not perfect and no ICA will be performed for mild stenosis to reveal the true underlying disease status, patients in this pathway can either be true positive, false negative (patients who actually suffer from 3VD/LM) or false positive (patients who actually don't suffer from CAD). Because it is recognized that some cardiologists will see the need for more aggressive treatment of mild-moderate stenosis, an alternative scenario was created in which 50% of patients with positive (but not markedly positive) tests or tests indicative of 1- or 2-vessel disease are sent directly for ICA, with rest receiving aggressive medical management. Similar to the base

case assumptions in the ED model, there is no empirical support for this distribution, but it was felt to be an important boundary for clinical decision-making by several members of the ERG (see modified diagnostic pathway in Appendix I).

Patients whose diagnostic test indicates no evidence of CAD receive no additional therapies beyond baseline care. Depending on the true disease status, they can either be true negative or false negative.

The 2-test strategy (Figure 3b) differs from the 1-test strategy in a way such that patients whose initial test is indeterminate or positive, but not markedly positive, for CAD will not immediately start on aggressive medical treatment nor be sent for ICA, but will receive a second test. The second test will then have three possible outcomes and resulting consequences that are identical to the 1-test strategy. Patients whose first test is either markedly positive for CAD or indicates no evidence of CAD will undergo no further testing and immediately receive the same management as outlined for the 1-test strategy. Under the aggressive management scenario, pathways for the *first* test are unchanged; patients with positive (but not markedly positive) results or results indicative of 1- or 2-vessel disease on the *second* test are managed as described above (see modified diagnostic pathway in Appendix I).

Figure 3: Diagnostic pathways



Notes: pos ++: markedly abnormal test result, pos +: abnormal test result, ind: indeterminate results: TP: true-positive; TN: true-negative; FP: false-positive; FN: false-negative; ICA: invasive coronary angiography; agg med mgmt: aggressive medical management (according to AHA/ACC guidelines)

Input Parameters

Clinical Parameters

Our base case cohort is 55 year old men with a CAD prevalence of 30% (intermediate prevalence). The proportion of patients among the different CAD severity levels was derived by averaging the data for 55 year old men with "non-anginal chest pain" and "atypical chest pain" as observed by Diamond and Forrester: 22% for one- or two-vessel CAD, 5% for three-vessel, and 3% for left main artery CAD (Diamond, 1979). When the overall CAD prevalence was modified to 10%, the ratio between the severity levels remained constant.

Test Accuracy

Test characteristics for CCTA were derived from our systematic review on a per-patient basis, and we assumed equal accuracy for one- or two vessel CAD and three-vessel or left main CAD. Note that the "as-reported" estimate for CCTA specificity (87%) was used rather than the "intent-to-diagnose" estimate (82%), as the model handles indeterminate findings separately. Test characteristics for stress-echocardiography and stress-SPECT were derived from published meta-analyses (Garber, 1999). All tests were considered to be conditionally independent (see Table IV below).

Table IV: Patient and diagnostic test characteristics

Variable	Base Case Estimate	Source(s)
Diagnostic test characteristics		
64-slice CCTA		
Sensitivity for CAD (per patient)	0.97	ICER Review
Specificity for CAD (per patient)	0.87	"
Indeterminate results	0.03	"
Stress ECHO		
Sensitivity for one- or two-vessel CAD	0.76	Garber, 1999
Sensitivity for three-vessel or left main CAD	0.94	"
Specificity for CAD	0.88	"
Indeterminate results	0.13	Ward, 2007
Stress SPECT		
Sensitivity for one- or two-vessel CAD	0.88	Garber, 1999
Sensitivity for three-vessel or left main CAD	0.98	"
Specificity for CAD	0.77	"
Indeterminate results	0.09	Patterson, 1995
ICA-related mortality	0.001	Kuntz, 1999

Notes: CCTA=coronary computed tomographic angiography; CAD=coronary artery disease; ECHO=echocardiogram; SPECT=single-photon emission computed tomography; ICA=invasive coronary angiography

Costs
Cost were estimated using Medicare reimbursement data (Centers for Medicare & Medicaid Services, 2008). Table V below depicts the detailed CPT codes associated with each cost item, including both the technical and the professional components for the reimbursement rate.

Table V: Cost estimates

Procedure, CPT code (description)		Total costs ($)	Source
SECHO	93015 (cardiovascular stress test) 93350 (echo transthoracic)	300	CMS, 2008
CCTA	0145T (CT heart w/wo dye funct: $306) Physician fee ($159)	466	"
SPECT	78465 (heart image (3d), multiple) 78478 (heart wall motion add-on) 78480 (heart function add-on) 93015 (cardiovascular stress test)	765	"
ICA	93508 (cath placement, angiography) 93510 (left heart catheterization) 93543 (injection for heart x-rays) 93545 (injection for coronary x-rays) 93555 (imaging, cardiac cath)	2,750	"

Notes: CCTA=coronary computed tomographic angiography; ECHO=echocardiogram; SPECT=single-photon emission computed tomography; ICA=invasive coronary angiography

Results

Base Case Analysis
Table VI on the following page depicts the results for 1,000 55-year old men with an underlying CAD prevalence of 30%. Each column represents the results if all patients had undergone the specific screening strategy.

From the data in Table VI on the following page it can be seen that there are important trade-offs to consider when comparing these strategies. For example, "*CCTA alone*" has the highest number of true positives at 288 and the lowest number of false negatives at 8 among all strategies, followed by "*SPECT alone*" which has 271 true positives and 25 false negatives; the number of false negatives with severe CAD (i.e., 3-vessel or left main disease) was not materially different between strategies, owing to the low actual prevalence of severe disease in this population. But CCTA strategies introduce the issue of incidental findings, estimated to require follow-up among 13.8% of all patients screened. CCTA (and

SPECT) strategies also carry radiation exposure risks for all patients. The strategy *"stress-ECHO followed by CCTA"* has the lowest cost per patient of $694 followed by *"CCTA alone"* with a cost of $760/patient. *"Stress-ECHO alone"* has the lowest number of patients exposed to any radiation with 195 due to invasive angiographies.

Table VI: Diagnostic results (30 % CAD prevalence)

Estimates	CCTA	SPECT	SECHO	CCTA -> SPECT	SPECT -> CCTA	CCTA -> SECHO	SECHO -> CCTA	SECHO -> SPECT
True positive	288	271	245	266	265	245	239	228
False positive	86	149	74	23	26	11	19	33
True negative	618	556	631	682	679	694	686	672
False negative	8	25	50	29	31	51	56	68
False negative w/3-v or LM disease	2	1	4	2	1	2	4	4
Referred for ICA	107	160	195	106	90	118	85	105
ICA-negative results	21	61	89	7	5	11	4	12
ICA related deaths	0.11	0.17	0.20	0.11	0.09	0.12	0.09	0.11
Exposed to radiation	1000	1000	195	1000	1000	1000	408	408
Incidental findings requiring f/u	138	0	0	138	57	138	47	0
Total costs/patient [excluding all f/u costs, $]	760	1,204	837	1,002	1,203	886	694	850

Notes: CCTA: coronary computed tomographic angiography; SPECT: single photon emission computed tomography; SECHO: stress echocardiogram; 3-v: 3-vessel coronary artery disease; LM: coronary artery disease of the left main artery; ICA: invasive coronary angiography; f/u: follow-up

When considering the outcomes and costs for this diagnostic phase only, *"CCTA alone"* is cost-saving and has fewer false negatives than all other strategies except *"stress-ECHO followed by CCTA,"* This latter two-test strategy is less costly and exposes less than half as many patients to radiation but also has more false negatives.

Because the general perception of the true underlying CAD prevalence associated with a "low-to-intermediate risk" population varies, we present Table VII on the following page depicting the result of the identical strategies for a population with 10% CAD prevalence. Comparing these results to table VI demonstrates the same ranking between the strategies with regard to accuracy, number of angiographies, number of incidental findings, and radiation exposure, thus resulting in the same interpretation. The lower diagnostic costs per

patient (11-26% reductions) compared to 30% CAD are primarily driven by the lower number of patients referred to ICA (30-62%).

Table VII : Diagnostic results (10% prevalence)

Estimates	CCTA	SPECT	SECHO	CCTA -> SPECT	SPECT -> CCTA	CCTA -> SECHO	SECHO -> CCTA	SECHO -> SPECT
True positive	96	91	82	89	89	81	80	75
False positive	111	190	94	29	33	15	25	43
True negative	790	711	807	872	868	887	876	858
False negative	3	8	17	10	10	18	19	24
False negative w/3-v or LM disease	1	0	1	1	0	1	1	1
Referred for ICA	56	111	151	41	35	49	32	46
ICA-negative results	28	78	116	11	7	16	5	17
ICA related deaths	0.06	0.11	0.15	0.04	0.04	0.05	0.03	0.05
Exposed to radiation	1000	1000	151	1000	1000	1000	292	292
Incidental findings requiring f/u	138	0	0	138	46	138	37	0
Total costs/patient [excluding all f/u costs, $]	619	1,071	714	740	1,017	663	514	634

Notes: CCTA: coronary computed tomographic angiography; SPECT: single photon emission computed tomography; SECHO: stress echocardiogram; 3-v: 3-vessel coronary artery disease; LM: coronary artery disease of the left main artery; ICA: invasive coronary angiography; f/u: follow-up

Aggressive Treatment of Mild-Moderate Stenosis

Under a scenario of more aggressive treatment for mild-moderate stenosis (50% ICA and 50% medical management), the number of people referred to invasive angiography doubled on average for both the 30% and the 10% prevalence group. As a consequence, the number of false positives decreased and the number of true negatives increased (as ICA always determines the true underlying disease status), while at the same time the number of ICA-related deaths nearly doubled and the price per patient increased by almost 30%.

In this version of the model, "*CCTA alone*" was more expensive than stress ECHO, stress ECHO followed by CCTA, CCTA followed by stress ECHO, and stress ECHO followed by

SPECT, as a consequence of a larger number of patients referred for ICA. For example, in the base case at 30% prevalence, the number of patients undergoing ICA in the CCTA alone and SECHO-SPECT strategies was essentially the same; in the more aggressive scenario, this number is increased by 30% for CCTA alone (see Appendix I for details).

Alternative Analysis of CCTA Test Performance

In recognition of the heterogeneity observed in the ICER meta-analysis of CCTA test characteristics, an alternative analysis was conducted using the sensitivity and specificity results from the CORE 64 multicenter study (Miller, 2008); these were 83% and 91% respectively, as compared to 97% and 87% in the model base case. Based on these estimates, the numbers of true and false positives declined in the CCTA-based strategies, while the number of false negatives increased to levels similar to those of the other strategies. Total costs were similar to the base case, however, owing to a lower number of patients referred for ICA; CCTA alone remained less costly than all other strategies except stress ECHO followed by CCTA (see Appendix I for details).

Lifetime Model

Survival

The basic approach taken to estimate the mortality risk ratios associated with one-, two-, three-vessel, and left main CAD was the development of a simulation model that predicted mortality in the COURAGE trial (Boden, 2007), generalizing the proportional relationship between risk ratios from a previous study (Kuntz, 1999). Specifically, survival was derived as a function of US life-tables stratified by age and gender and risk ratios accounting for the number of diseased vessels (1.4 for one- or two-vessel CAD, 2.2. for 3-vessel and 5.8 for left main artery disease). Lack of appropriate treatment (PCI or meds for one- or two-vessel CAD, PCI and meds for three-vessel CAD, PCI and CABG for left main CAD) increased mortality risk by an additional 30% (LaRosa, 1999). Note that CAD-negative patients could subsequently develop CAD and the disease could progress.

Utilities

Utilities were also derived from the COURAGE trial (Boden, 2007) and depended on whether the patient had no CAD (0.96), CAD without chest pain (0.88) or CAD with chest pain (0.78). Occluded arteries caused chest pain; appropriate treatment relieved chest pain, resulting in a pain-free fraction after one year of 74% for CABG (Hoffman, 2003), 66% for PCI (Boden, 2007), 58% for medical treatment (Boden, 2007), and 13% in patients without treatment (Boden, 2007).

Costs

In addition to the one-time cost for the diagnostic work-up (Table IV), additional costs were accounted for as they occurred. PCI and CABG were assigned costs of $11,210 (Cohen, 2004) and $25,500, respectively (Reynolds, 2003). In addition, all patients received baseline prophylaxis consisting of Aspirin (81 mg QD) and simvastatin (20 mg QD) at $310/year (Drugstore.com, 2007). Patients who suffered from chest pain also received symptomatic treatment for angina consisting of atenolol (50 mg QD) and isosorbide mononitrate (60 mg

QD) assigned a cost of $170/year (Drugstore.com, 2007). Note that, due to time constraints, downstream costs due to cardiac events in false-negative patients (e.g., missed MI) were not included; such patients do receive a QALY decrement, however, from both an increased mortality risk from inappropriate treatment and from untreated chest pain.

Effects of Diagnostic Accuracy

The effect of the different diagnostic strategies is modeled indirectly via the proportion of patients correctly and incorrectly classified with respect to CAD status and resulting treatment action. True positives are assumed to be treated accordingly, thus profiting from a survival and quality of life benefit while true negatives do not undergo an invasive angiography and thus do not experience the risk of intervention-related mortality and costs. False negatives do not profit from the treatment appropriate for their severity of disease and thus experience no benefit in survival and quality-of-life as compared to those who are treated appropriately. Lastly, a small portion of false positives will die from unnecessarily performed ICA and all false positives will generate costs due unnecessary treatment.

Results

Base Case Analysis: CAD Prevalence 30%

Table VIII below depicts the remaining quality adjusted life years (QALY) and lifetime medical cost as predicted for the different strategies for 55 year old men with a CAD prevalence of 30%. Note that the QALY range between the most effective and least effective strategy is only 16 days. This small difference appears very reasonable as the diagnostic test is a one-time evaluation. The dynamic nature of the model is built to reflect clinical reality, allowing for initially healthy patients to develop disease over time and for CAD to progress, both situations that will require future treatment and revascularization.

Table VIII: Strategies ordered by increasing effectiveness (30% CAD prevalence)

Strategy	Effectiveness (QALY)	Costs ($)
SECHO-SPECT	15.140	7,576
CCTA-SECHO	15.146	7,605
SECHO-CCTA	15.151	7,343
CCTA-SPECT	15.154	7,911.
SPECT-CCTA	15.157	8,077
SECHO	15.167	7,998
SPECT	15.172	9,051
CCTA	15.183	8,207

Notes: QALY: quality-adjusted life year

Incremental Cost-Effectiveness Analysis

Note that, while all strategies are included for informational purposes, incremental cost-effectiveness results are calculated where feasible for the following comparisons of primary interest: CCTA alone vs. Stress ECHO alone and vs. SPECT alone, as well as for the least expensive and most effective strategies involving CCTA relative to Stress ECHO alone. For 30% CAD prevalence, "*CCTA alone*" is the most effective strategy, while "*Stress ECHO followed by CCTA*" is the least expensive.

Comparing CCTA and Stress ECHO (Table IX), "*CCTA alone*" results in a gain of an additional 0.016 QALYs and comes at an additional cost of $209, which can be converted into an incremental cost-effectiveness ratio of about $13,100/QALY. An incremental cost-effectiveness ratio cannot be generated for CCTA vs. SPECT, as CCTA is both more effective and less expensive and thus dominates SPECT. "*Stress ECHO followed by CCTA*" is both less expensive and less effective than Stress ECHO alone, and an incremental cost-effectiveness ratio is not generated.

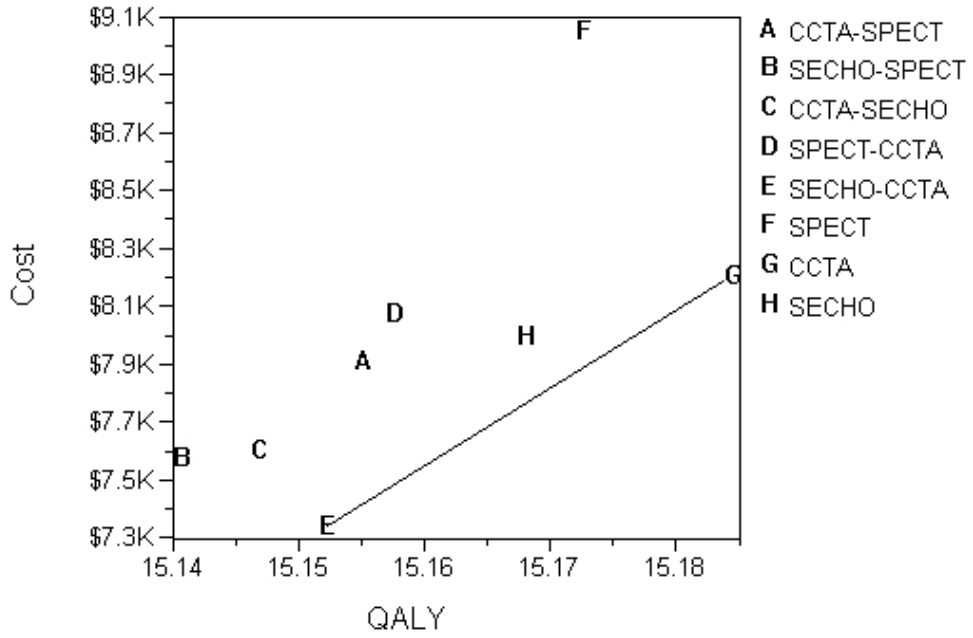
Table IX: Incremental cost-effectiveness analysis (30% CAD prevalence)

Strategy	Effect	Incr. Effect	Costs	Incr. Costs	Incr. C/E
SECHO	15.167		7,998		
CCTA	15.183	0.016	8,207	209	13,100

Notes: CCTA: Computed Coronary Tomography Angiography, SECHO: Stress Echocardiogram, SPECT: Single Photon Emission Computed Tomography

For informational purposes, Figure 4 on the following page depicts the results of all strategies graphically. The y-axis shows the life-time medical costs and the x-axis the quality-adjusted life gain associated with each strategy. "Stress Echo followed by CCTA" is the least expensive strategy and thus the reference. The line between "Stress-echo followed by CCTA" [E] and "CCTA alone" [G] shows the cost-effectiveness frontier; all strategies above this frontier are dominated.

Figure 4: Cost-effectiveness graph (30% CAD prevalence)



Notes: CCTA: Computed Coronary Tomography Angiography, SECHO: Stress Echocardiogram, SPECT: Single Photon Emission Computed Tomography

When a more aggressive treatment for mild-moderate stenosis (50% ICA and 50% medical management) is considered, the average life-expectancy for CCTA increases slightly by 0.7 quality adjusted life days (15.185 QALY vs. 15.183 QALY) and this increase comes at additional cost of $595 ($ 8,802 vs. $8,207). Generally, the life-expectancies for all strategies in this scenario increase slightly as compared to 100% medical management. However, the strategies with the lowest effectiveness [combination strategies] improve the most, and while the ranking is preserved for the most part, the overall range between the least and the most effective strategy decreases to about 9 quality adjusted life days. Lifetime costs increase between $400 and $600. Comparing CCTA to stress-echo alone results in incremental cost-effectiveness ratio of $16,100/ QALY (for further detail, see Appendix G).

Base Case Analysis: CAD Prevalence 10%

Table X on the following page depicts the remaining quality adjusted life years (QALY) and lifetime medical cost as predicted for the different strategies for 55 year old men with a CAD prevalence of 10%. Note that for a CAD prevalence of 10%, the difference in QALYs between the most and the least effective strategy decreases to 7 days.

Table X: Strategies ordered by increasing effectiveness (10% CAD prevalence)

Strategy	Effectiveness (QALY)	Costs ($)
SECHO	16.012	4,543
CCTA-SECHO	16.014	3,962
SECHO-SPECT	16.014	4,068
SECHO-CCTA	16.015	3,831
CCTA-SPECT	16.017	4,175
CCTA	16.018	4,645
SPECT-CCTA	16.024	4,450
SPECT	16.030	5,633

Notes: CCTA: Computed Coronary Tomography Angiography, SECHO: Stress Echocardiogram, SPECT: Single Photon Emission Computed Tomography

Incremental Cost-Effectiveness Analysis

As in the case of 30% CAD prevalence, strategies were compared to a referent category of "*Stress ECHO alone*", and incremental cost-effectiveness ratios are generated for *CCTA* vs. *Stress ECHO* and *SPECT*, as well as for the most effective strategy involving CCTA ("*SPECT followed by CCTA*") and the least expensive strategy involving CCTA ("*Stress ECHO followed by CCTA*".

It is important to note the implications of the changes in cost-effectiveness results between the 30% and the 10% prevalence populations. As the prevalence of CAD in the tested population goes lower, the risk of false-negative results is diminished, whereas the risk of false-positive results is increased. This shift will tend to enhance the diagnostic utility of strategies with lower sensitivity and higher specificity relative to other strategies. Thus, in comparison to the results for the 30% prevalence population, the results for the 10% prevalence population are driven much more by the false-positive rate than by the false-negative rate. If the CAD prevalence in the tested population drops lower than 10%, the incremental cost-effectiveness ratios for CCTA and CCTA-based strategies will continue to rise in comparison to Stress ECHO.

Comparing CCTA vs. Stress ECHO (Table XI on the following page), "*CCTA alone*" gains an additional 0.006 QALYs at an incremental cost of $102, for an incremental cost-effectiveness ratio of $17,000/QALY.

Table XI: Incremental cost-effectiveness analysis (10% CAD prevalence)

Strategy	Effect	Incr. Effect	Costs	Incr. Costs	Incr. C/E
SECHO	16.012		4,543		
CCTA	16.018	0.006	4,645	102	17,000

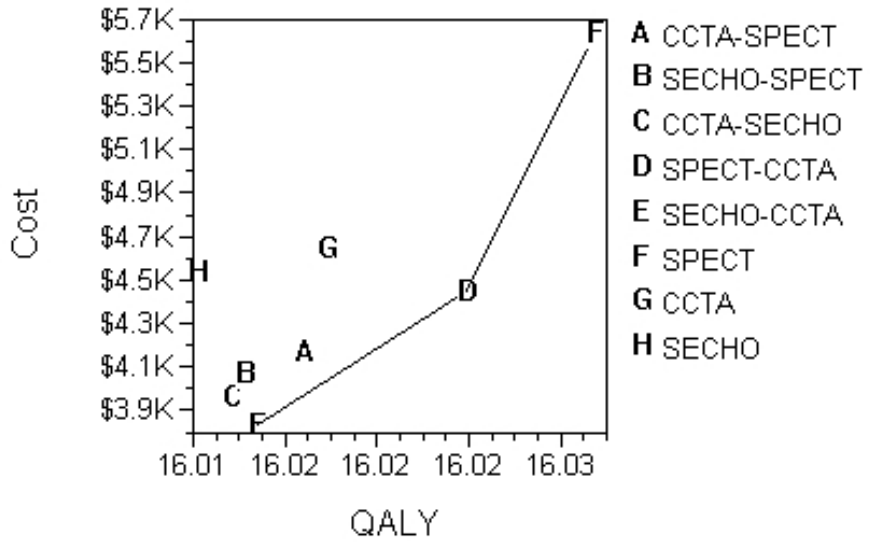
When comparing single-test strategies involving CCTA and SPECT (Table XII), findings contrast with the 30% results in that CCTA is now less effective than SPECT. Because SPECT is also more expensive, an incremental cost-effectiveness ratio of $82,300 is generated for SPECT vs. CCTA. When comparing *"SPECT followed by CCTA"* (the most effective strategy involving CCTA), and *"stress ECHO followed by CCTA"* (the least expensive strategy involving CCTA), both strategies were more effective than Stress ECHO alone and less costly, so incremental cost-effectiveness ratios were not generated.

Table XII: Incremental cost-effectiveness analysis (CCTA vs. SPECT)

Strategy	Effect	Incr. Effect	Costs	Incr. Costs	Incr. C/E
CCTA	16.018		4,645		
SPECT	16.030	0.012	5,633	988	82,300

Figure 5 on the following page depicts the results graphically for all strategies at 10% CAD prevalence. The y-axis shows the life-time medical costs and the x-axis the quality-adjusted life gained associated with each strategy. The line between "*stress-echo followed by CCTA*" [E], "*SPECT followed by CCTA*" [D] and "*SPECT alone*" [E] depicts the cost-effectiveness frontier; all strategies above this frontier are dominated.

Figure 5: Cost-effectiveness graph (10% CAD prevalence)



Notes: CCTA: Computed Coronary Tomography Angiography, SECHO: Stress Echocardiogram, SPECT: Single Photon Emission Computed Tomography

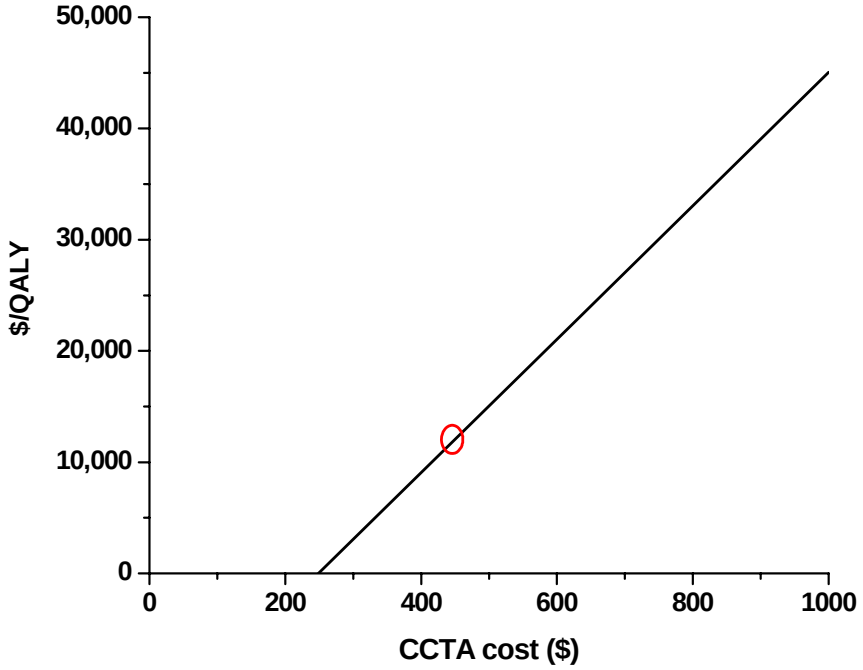
When a more aggressive treatment for mild-moderate stenosis (50% ICA and 50% medical management) is considered, the overall life expectancy for all strategies decreases slightly (~ 7 quality adjusted life days) compared to 100% medical management as a consequence of a greater number of ICA-related deaths. Costs decrease on average by about $500 for each strategy, as a consequence of lower expenses due to lower rates of inappropriately initiated medical treatment. Because of its reduced specificity compared to CCTA (0.77 vs. 0.87), the outcome for "*SPECT*" is affected in such a way that it is now dominated by "*CCTA*". Comparing "*CCTA*" to "*stress Echo alone*" results in an incremental cost-effectiveness ratio of $12,700/ QALY (for further detail, see Appendix I).

Sensitivity Analyses (30% CAD Prevalence)

<u>CCTA Costs</u>

CCTA costs occur as a one-time cost for those patients who underwent CCTA as part of their diagnostic work-up. For the base case we assumed a cost of $466, resulting in an incremental cost-effectiveness ratio (ICER) of about $13,100/QALY. Figure 6 on the following page depicts the linear relationship between CCTA costs and the ICER comparing "*CCTA alone*" to "*stress-echo alone*". For a CCTA cost of about $248 or less, CCTA is dominant.

Figure 6: Sensitivity analysis: CCTA cost (30% CAD prevalence)



Note: circle: base case estimate for CCTA cost.

CCTA Test Performance

As described for the diagnostic phase above, we examined the impact on our lifetime results if the CORE 64 estimates of CCTA's diagnostic accuracy were used (Miller, 2008). The resulting decrease in both true and false positives leads to reduced lifetime costs, as fewer patients are referred for ICA and receive CABG or PCI, and fewer false positives incur the costs of drug therapy. In addition, the increase in false negatives leads to a greater number of patients treated inappropriately, reducing CCTA's quality adjusted life expectancy from 15.183 to 15.176 QALYs. However, CCTA alone remains the most effective of the 8 strategies, and because its cost is reduced from $8,207 in the base case to $7,581, it is both more effective and less costly than either stress ECHO or SPECT alone (see Appendix I for details).

Model Considerations and Limitations

As with all decision analytic and cost-utility models, our models required many assumptions and judgments. Among these, it is important to note again that all analyses were performed without considering harm, benefit, or costs of radiation-exposure or incidental findings. *"CCTA alone"* resulted in about 14% incidental findings and thus required follow-up as compared to 0-5% in the other strategies. Strategies including either CCTA or SPECT as the first or only test exposed all patients to radiation, as opposed to 20-40% of patients exposed in strategies with stress-ECHO as the first or only test.

One aspect of the models that should also be noted is the way that the health impact of a "false positive" was modeled. While false negatives in the model experience a negative health outcome due to lack of appropriate treatment (although there is no financial cost assigned to a false-negative diagnosis per se), there is no negative health impact of a false positive diagnosis; the model only accounts for the unnecessary health care costs for false-positives. Indeed, in the lifetime model some of the false positives develop CAD during the course of the simulation, in which case they would later profit from the initially unnecessary treatment.

In addition, the model assumed conditional independence of test performance for both the single-test and dual-test strategies. In reality, the results of one test will likely complement the interpretation of the second test by its impact on the pretest probability of disease (Kroenke, 1992); results of each of the two tests will be viewed in combination rather than in isolation. Due to the complexity of such a modeling approach, test "complementarity" was not examined, and effectiveness of the dual-test strategies may have been underestimated as a result.

Conclusions

At a CAD prevalence of 30%, CCTA produces a higher number of true positives and fewer false negatives relative to other 1- or 2-test strategies, and lower diagnostic phase costs than nearly all other tests; at a prevalence of 10%, differences in test performance are diminished but the pattern of costs remains the same. When alternative estimates of CCTA's diagnostic accuracy are employed, the balance of false-positive and false-negative shifts, but has little impact on comparative cost between the strategies. However, when a more aggressive strategy for management of mild-moderate stenosis is employed, CCTA becomes more costly than several other strategies due to a higher rate of referral for ICA.

Considering a lifetime horizon, quality-adjusted life expectancy is quite similar across the strategies, with a difference of only about 2 weeks between the most and least effective strategies. At 30% CAD prevalence, a single-test strategy with CCTA appears to be more effective and less costly than SPECT, and a reasonable value when compared to Stress ECHO (incremental cost-effectiveness ratios of $13,000-$16,000/QALY). When prevalence is reduced to 10%, however, while cost-effectiveness is similar for CCTA vs. Stress ECHO, SPECT is more effective than CCTA at a ratio of approximately $80,000/QALY. A shift from conservative to aggressive management of mild-moderate stenosis affects the lifetime results only marginally, as does the use of alternative estimates of CCTA's diagnostic accuracy.

Because the range of effectiveness results is so narrow, the model is highly sensitive to changes in selected parameters, in particular the costs of the various strategies. For example, at a cost of $248 or less, CCTA would dominate all other strategies, while for CCTA costs of $1,083, $1,916, and $2,749, the cost-effectiveness ratios would be $50,000/QALY, $100,000/QALY, and $150,000/QALY, respectively.

9. Recommendations for Future Research

As documented in this appraisal, there are numerous remaining areas of uncertainty regarding the impact on patient outcomes and resource utilization of CCTA in the ED and outpatient settings. Based on an assessment of which future research findings would have the greatest impact on judgments of the comparative clinical effectiveness and value of CCTA, ICER recommends that studies be pursued to address the following questions:

1) Do differences in test performance between CCTA and other non-invasive diagnostic strategies translate into clinical outcome differences?

As noted in this review, there is relatively consistent evidence on the sensitivity and specificity of CCTA in comparison to ICA. However, CCTA provides a visual analogue to ICA results, whereas stress ECHO and SPECT provide clinicians with functional information. Since the non-invasive alternatives all give different information to clinicians, it is very difficult to judge whether the higher sensitivity of CCTA identifies patients with CAD who will benefit from treatment to the same extent as patients identified through other means. The lack of published evidence on the impact of CCTA on clinician decision-making, rates of invasive angiography, and subsequent major cardiac events thus represents a particularly important evidence gap. Particularly for outpatient evaluation, randomized controlled trials are needed. In the ED setting, randomization would ideally occur following an initial negative serum enzyme and negative or non-specific EKG finding; in the outpatient setting best evidence would result if randomization were possible at the time patients are first considered for an outpatient non-invasive study. Several multi-center randomized trials are currently under consideration for funding by the National Institutes of Health (personal communication, Pamela Douglas, MD, December 5, 2008). In order to address evidence gaps most effectively, these RCTs should ideally include clinicians in community settings and enroll patients with few exclusions. Data should be gathered on the impact of CCTA on subsequent testing and treatment decisions as well as on major cardiac events, requiring a duration of follow-up of at least one year.

Other prospective studies could complement longer RCTs by focusing on the impact of CCTA on immediate triage and treatment decisions; an example of a study design well conceived to accomplish this is that used by Rubinshtein (Rubinshtein, 2007 [3]) of triage decisions in the ED. By allowing physicians to indicate an initial triage decision before providing CCTA results, the Rubinshtein article provides excellent quality evidence of the potential impact of CCTA on decision-making. Multiple studies of this type from different practice locations would be useful to enhance the generalizability of the findings.

2) What is the long-term impact of CCTA on poorly studied outcomes including incidental findings, secondary malignancies, and longer-term re-testing outcomes in low-to-intermediate risk populations?

Many of the unknowns regarding CCTA relate to outcomes that can only be reliably assessed with extensive follow-up. These include the effects of radiation dose on the incidence of secondary cancers, the rate incidental findings and the clinical and economic outcomes associated with their follow-up, and long-term outcomes for patients treated medically or invasively for CAD based on CCTA findings. A registry would likely be the ideal vehicle for such an evaluation, given the need for long-term follow-up and detailed clinical information. However, given the need to compare how patients treated based on CCTA results fare relative to those diagnosed via other means, an expansion of the registry construct to include other means of diagnosing CAD might be warranted.

3) *Will widespread availability of CCTA change the clinician's and/or patient's threshold for testing?*

One of the most vexing unknowns about introduction of a new diagnostic test is whether its introduction will lower the general threshold for testing. From the patient perspective, CCTA has some attractive features – it is relatively quick to perform compared to the functional tests, it is relatively painless, and there is no need to exercise – which may in turn lead to increased demand for the test among patients previously thought to be at too low a risk for CAD diagnostic testing. In addition, while use of CCTA for detection of early disease in asymptomatic individuals is not recommended by current guidelines, some clinicians may find this information important enough to warrant testing.

The best study design for the purpose of assessing the impact of CCTA on testing thresholds would involve serial population-based measures of CCTA use, other non-invasive testing, and ICA rates in a population for which CCTA is available vs. the same measures in a control population for which CCTA is not available. Age, sex, and co-morbidity adjusted rates/1,000 of these outcomes should help elucidate the degree to which CCTA availability lowers the general testing threshold, and whether any increased overall non-invasive testing leads to higher population-based rates of invasive ICA.

REFERENCES

Achenbach S, Ropers U, Kuettner A et al. Randomized Comparison of 64-Slice Single- and Dual-Source Computed Tomography Coronary Angiography for the Detection of Coronary Artery Disease. *J Am Coll Cardiol* 2008; 1 (2):177-86.

American Heart Association. Heart Disease and Stroke Statistics – 2008 Update. Dallas, Texas: American Heart Association. http://circ.ahajournals.org/cgi/reprint/CIRCULATIONAHA.107.187998 2008. (accessed 2008).

Ambrose JA. In search of the "vulnerable plaque": can it be localized and will focal regional therapy ever be an option for cardiac prevention? *J Am Coll Cardiol* 2008; 51 (16):1539-42.

Anthony D. Diagnosis and screening of coronary artery disease. *Prim Care* 2005; 32 (4):931-46.

Bairey Merz CN, Shaw LJ, Reis SE et al. Insights from the NHLBI-Sponsored Women's Ischemia Syndrome Evaluation (WISE) Study: Part II: gender differences in presentation, diagnosis, and outcome with regard to gender-based pathophysiology of atherosclerosis and macrovascular and microvascular coronary disease. *J Am Coll Cardiol* 2006; 47 (3 Suppl):S21-9.

Bayrak F, Guneysu T, Gemici G et al. Diagnostic performance of 64-slice computed tomography coronary angiography to detect significant coronary artery stenosis. *Acta Cardiol* 2008; 63 (1):11-7.

Boden WE, O'Rourke RA, Teo KK, et al. Optimal Medical Therapy with or without PCI for Stable Coronary Disease. *N Engl J Med* 2007;356(15):1503-16.

Brenner DJ, Georgsson MA. Mass screening with CT colonography: should the radiation exposure be of concern? *Gastroenterology* 2005; 129 (1):328-37.

Budoff MJ, Achenbach S, Blumenthal RS et al. Assessment of coronary artery disease by cardiac computed tomography: a scientific statement from the American Heart Association Committee on Cardiovascular Imaging and Intervention, Council on Cardiovascular Radiology and Intervention, and Committee on Cardiac Imaging, Council on Clinical Cardiology. *Circulation* 2006; 114 (16):1761-91.

Budoff MJ, Cohen MC, Garcia MJ et al. ACCF/AHA clinical competence statement on cardiac imaging with computed tomography and magnetic resonance: a report of the American College of Cardiology Foundation/American Heart Association/American College of Physicians Task Force on Clinical Competence and Training. *J Am Coll Cardiol* 2005; 46 (2):383-402.

Budoff MJ, Dowe D, Jollis JG et al. Diagnostic Performance of 64-Multidetector Row Coronary Computed Tomographic Angiography for Evaluation of Coronary Artery Stenosis in Individuals Without Known Coronary Artery Disease: Results From the Prospective Multicenter ACCURACY (Assessment by Coronary Computed Tomographic Angiography of Individuals Undergoing Invasive Coronary Angiography) Trial. *J Am Coll Cardiol* 2008; 52:000-.

Cademartiri F, Maffei E, Notarangelo F et al. [1] 64-slice computed tomography coronary angiography: diagnostic accuracy in the real world. *Radiol Med* 2008; 113 (2):163-80.

Cademartiri F, Maffei E, Palumbo A et al. [2] Diagnostic accuracy of 64-slice computed tomography coronary angiography in patients with low-to-intermediate risk. *Radiol Med* 2007; 112 (7):969-81.

Cademartiri F, Maffei E, Palumbo AA et al. [3] Influence of intra-coronary enhancement on diagnostic accuracy with 64-slice CT coronary angiography. *Eur Radiol* 2008; 18 (3):576-83.

Cademartiri F, Malago R, Belgrano M et al. [4] Spectrum of collateral findings in multislice CT coronary angiography. *Radiol Med* (Torino) 2007; 112 (7):937-48.

Caro JJ, Trindade E, McGregor M. The risks of death and of severe nonfatal reactions with high- vs low-osmolality contrast media: a meta-analysis. *AJR Am J Roentgenol* 1991; 156 (4):825-32.

California Technology Assessment Forum. Computed Tomographic Angiography in the Diagnosis of Coronary Artery Stenosis and for the Evaluation of Acute Chest Pain. http://www.ctaf.org/content/general/detail/768. 2007 (accessed 2008).

Centers for Disease Control and Prevention. Heart Disease Facts and Statistics. http://www.cdc.gov/heartdisease/facts.htm. 2007. (accessed Aug 19, 2008).

Centers for Medicare & Medicaid Services. CMS Clinical Lab Fee Schedule Website. http://www.cmshhsgov/clinicalfeesched/. (accessed Sep 03, 2008).

Chaitman BR, Bourassa MG, Davis K, et al. Angiographic prevalence of high-risk coronary artery disease in patient subsets (CASS). *Circulation* 1981;64(2):360-7.

Chang AM, Shofer FS, Weiner MG, et al. Actual financial comparison of four strategies to evaluate patients with potential acute coronary syndromes. *Acad Emerg Med* 2008; 15 (7):649-55.

Cohen DJ, Bakhai A, Shi C, et al. Cost-effectiveness of sirolimus-eluting stents for treatment of complex coronary stenoses: results from the Sirolimus-Eluting Balloon Expandable Stent in the Treatment of Patients With De Novo Native Coronary Artery Lesions (SIRIUS) trial. *Circulation* 2004;110(5):508-14.

Danciu SC, Herrera CJ, Stecy PJ, et al. Usefulness of multislice computed tomographic coronary angiography to identify patients with abnormal myocardial perfusion stress in whom diagnostic catheterization may be safely avoided. *Am J Cardiol* 2007; 100 (11):1605-8.

Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* 2005; 58 (9):882-93.

DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7 (3):177-88.

Dewey M, Schnapauff D, Teige F, et al. Non-cardiac findings on coronary computed tomography and magnetic resonance imaging. *Eur Radiol* 2007; 17 (8):2038-43.

Diamond GA, Forrester JS. Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease. *N Engl J Med* 1979; 300 (24):1350-8.

Drugstore.com Website. http://wwwdrugstorecom. 2007. (accessed July 01,2007).

Earls JP, Berman EL, Urban BA, et al. Prospectively gated transverse coronary CT angiography versus retrospectively gated helical technique: improved image quality and reduced radiation dose. *Radiology* 2008; 246 (3):742-53.

Ehara M, Surmely JF, Kawai M, et al. Diagnostic accuracy of 64-slice computed tomography for detecting angiographically significant coronary artery stenosis in an unselected consecutive patient population: comparison with conventional invasive angiography. *Circulation* 2006; 70 (5):564-71.

Einstein AJ, Henzlova MJ, Rajagopalan S. Estimating risk of cancer associated with radiation exposure from 64-slice computed tomography coronary angiography. *JAMA* 2007; 298 (3):317-23.

El-Hajjar M, Bashir I, Khan M et al. Incidence of contrast-induced nephropathy in patients with chronic renal insufficiency undergoing multidetector computed tomographic angiography treated with preventive measures. *Am J Cardiol* 2008; 102 (3):353-6.

FDA. What are the Radiation Risks from CT? http://www.fda.gov/cdrh/ct/risks.html 2008. (accessed 2008).

Fine JJ, Hopkins CB, Ruff N, et al. Comparison of accuracy of 64-slice cardiovascular computed tomography with coronary angiography in patients with suspected coronary artery disease. *Am J Cardiol* 2006; 97 (2):173-4.

Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991; 11 (2):88-94.

Gaemperli O, Schepis T, Koepfli P, et al. Accuracy of 64-slice CT angiography for the detection of functionally relevant coronary stenoses as assessed with myocardial perfusion SPECT. *Eur J Nucl Med Mol Imaging* 2007; 34 (8):1162-71.

Gallagher MJ, Ross MA, Raff GL, et al. The diagnostic accuracy of 64-slice computed tomography coronary angiography compared with stress nuclear imaging in emergency department low-risk chest pain patients. *Ann Emerg Med* 2007; 49 (2):125-36.

Gandelman G. Medical Encyclopedia: Coronary angiography. http://www.nlm.nih.gov/medlineplus/ency/article/003876.htm 2006 (accessed Sep 4, 2008).

Garber AM, Solomon NA. Cost-effectiveness of alternative test strategies for the diagnosis of coronary artery disease. *Ann Intern Med* 1999;130(9):719-28.

Gedik GK, Ergun EL, Aslan M, et al. Unusual extracardiac findings detected on myocardial perfusion single photon emission computed tomography studies with Tc-99m sestamibi. *Clin Nucl Med* 2007; 32 (12):920-6.

Gerson DA, Rybicki FJ, Yucel EK et al. Chronic chest pain - suspected cardiac origin. Appropriateness Criteria (R). Series Chronic chest pain - suspected cardiac origin. Appropriateness Criteria (R). http://www.acr.org 2006. (accessed 2008).

Ghostine S, Caussin C, Daoud B, et al. Non-invasive detection of coronary artery disease in patients with left bundle branch block using 64-slice computed tomography.[see comment]. *J Am Coll Cardiol* 2006; 48 (10):1929-34.

Gibbons E. Less invasive heart tests. *Health News* 2002; 8 (3):1-2, 9.

Gibbons RJ, Abrams J, Chatterjee K, et al. ACC/AHA 2002 guideline update for the management of patients with chronic stable angina--summary article: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on the Management of Patients With Chronic Stable Angina). *Circulation* 2003; 107 (1):149-58.

Gil BN, Ran K, Tamar G, et al. Prevalence of significant noncardiac findings on coronary multidetector computed tomography angiography in asymptomatic patients. *J Comput Assist Tomogr* 2007; 31 (1):1-4.

Goehring C, Perrier A, Morabia A. Spectrum bias: a quantitative and graphical analysis of the variability of medical diagnostic test performance. *Stat Med* 2004; 23 (1):125-35.

Gold MR, Siegel JR, Russell LB, Weinstein MC. Cost-effectiveness in health and medicine. New York: Oxford University Press. 1996.

Goldman L, Cook EF, Brand DA, et al. A computer protocol to predict myocardial infarction in emergency department patients with chest pain. *N Engl J Med* 1988;318(13):797-803.

Goldstein JA, Gallagher MJ, O'Neill WW, et al. A randomized controlled trial of multi-slice coronary computed tomography for evaluation of acute chest pain.[see comment]. *J Am Coll Cardiol* 2007; 49 (8):863-71.

Hacker M, Jakobs T, Hack N, et al. Sixty-four slice spiral CT angiography does not predict the functional relevance of coronary artery stenoses in patients with stable angina.[see comment]. *Eur J Nucl Med Mol Imaging* 2007; 34 (1):4-10.

Hausleiter J, Meyer T, Hermann F et al. International PRospective Multicenter Study on RadiaTion dose Estimates Of Cardiac CT AngIOgraphy IN Daily Practice. In American College of Cardiology 57th Annual Scientific Session, March 31, 2008. Chicago, IL; 2008.

Healthcare Human Factors Group. Computed Tomography Radiation Safety: Issues in Ontario 2008. Center for Global eHealth Innovation. http://www.ehealthinnovation.org/?q=node/363. 2006. (accessed 2008).

Hendel RC, Patel MR, Kramer CM, et al. ACCF/ACR/SCCT/SCMR/ASNC/NASCI/SCAI/SIR 2006 appropriateness criteria for cardiac computed tomography and cardiac magnetic resonance imaging: a report of the American College of Cardiology Foundation Quality Strategic Directions Committee Appropriateness Criteria Working Group, American College of Radiology, Society of Cardiovascular Computed Tomography, Society for Cardiovascular Magnetic Resonance, American Society of Nuclear Cardiology, North American Society for Cardiac Imaging, Society for Cardiovascular Angiography and Interventions, and Society of Interventional Radiology. *J Am Coll Cardiol* 2006; 48 (7):1475-97.

Ho JS, Fitzgerald SJ, Stolfus LL et al. Relation of a coronary artery calcium score higher than 400 to coronary stenoses detected using multidetector computed tomography and to traditional cardiovascular risk factors. *Am J Cardiol* 2008; 101 (10):1444-7.

Hoffmann U, Ferencik M, Cury RC et al. Coronary CT angiography. *J Nucl Med* 2006; 47 (5):797-806.

Hoffmann U, Nagurney JT, Moselewski F, et al. Coronary multidetector computed tomography in the assessment of patients with acute chest pain.[see comment][erratum appears in Circulation. 2006 Dec 19;114(25):e651]. *Circulation* 2006; 114 (21):2251-60.

Hoffman SN, TenBrook JA, Wolf MP, et al. A meta-analysis of randomized controlled trials comparing coronary artery bypass graft with percutaneous transluminal coronary angioplasty: one- to eight-year outcomes. *J Am Coll Cardiol* 2003;41(8):1293-304.

Hollander JE, Litt HI, Chase M, et al. Computed tomography coronary angiography for rapid disposition of low-risk emergency department patients with chest pain syndromes. *Acad Emerg Med* 2007; 14 (2):112-6.

Husmann L, Schepis T, Scheffel H, et al. [1] Comparison of diagnostic accuracy of 64-slice computed tomography coronary angiography in patients with low, intermediate, and high cardiovascular risk. *Acad Radiol* 2008; 15 (4):452-61.

Husmann L, Valenta I, Gaemperli O, et al. [2] Feasibility of low-dose coronary CT angiography: first experience with prospective ECG-gating. *Eur Heart J* 2008; 29 (2):191-7.

Jacobs JE, Boxt LM, Desjardins B, et al. ACR practice guideline for the performance and interpretation of cardiac computed tomography (CT). *J Am Coll Radiol* 2006; 3 (9):677-85.

Jeetley P, Hickman M, Kamp O, et al. Myocardial contrast echocardiography for the detection of coronary artery stenosis: a prospective multicenter study in comparison with single-photon emission computed tomography. *J Am Coll Cardiol* 2006; 47 (1):141-5.

Johnson TR, Nikolaou K, Busch S, et al. [2] Diagnostic accuracy of dual-source computed tomography in the diagnosis of coronary artery disease. *Invest Radiol* 2007; 42 (10):684-91.

Johnson TR, Nikolaou K, Wintersperger BJ, et al. [1] ECG-gated 64-MDCT angiography in the differential diagnosis of acute chest pain. *AJR Am J Roentgenol* 2007; 188 (1):76-82.

Khare RK, Courtney DM, Powell ES, et al. Sixty-four-slice computed tomography of the coronary arteries: cost-effectiveness analysis of patients presenting to the emergency department with low-risk chest pain. *Acad Emerg Med* 2008; 15 (7):623-32.

Kirsch J, Araoz PA, Steinberg FB, et al. Prevalence and significance of incidental extracardiac findings at 64-multidetector coronary CTA. *J Thorac Imaging* 2007; 22 (4):330-4.

Klocke FJ, Baird MG, Lorell BH, et al. ACC/AHA/ASNC guidelines for the clinical use of cardiac radionuclide imaging--executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (ACC/AHA/ASNC Committee to Revise the 1995 Guidelines for the Clinical Use of Cardiac Radionuclide Imaging). *J Am Coll Cardiol* 2003; 42 (7):1318-33.

Kroenke K, Lucas CA, Rosenberg ML, et al. Causes of persistent dizziness: a prospective study of 100 patients in ambulatory care. *Ann Intern Med* 1992; 117 (11):898-904.

Kuntz KM, Fleischmann KE, Hunink MG, et al. Cost-effectiveness of diagnostic strategies for patients with chest pain. *Ann Intern Med* 1999;130(9):709-18.

Ladapo JA, Hoffmann U, Bamberg F, et al. Cost-effectiveness of coronary MDCT in the triage of patients with acute chest pain. *Am J Roentgenol* 2008; 191(2):455-63.

LaRosa JC, He J, Vupputuri S. Effect of Statins on Risk of Coronary Disease: A Meta-analysis of Randomized Controlled Trials. *JAMA* 1999;282(24):2340-6.

Lau J, Ioannidis JP, Balk EM et al. Diagnosing acute cardiac ischemia in the emergency department: a systematic review of the accuracy and clinical effect of current technologies. *Ann Emerg Med* 2001; 37 (5):453-60.

Lau J, Kent D, Tatsioni A et al. Vulnerable Plaques: A Brief Review of the Concept and Proposed Approaches to Diagnosis and Treatment, Technology Assessment. In., Series Vulnerable Plaques: A Brief Review of the Concept and Proposed Approaches to Diagnosis and Treatment, Technology Assessment. Rockville, MD: Agency for Healthcare Research and Quality; 2004.

Leber AW, Johnson T, Becker A, et al. Diagnostic accuracy of dual-source multi-slice CT-coronary angiography in patients with an intermediate pretest likelihood for coronary artery disease.[see comment]. *Eur Heart J* 2007; 28 (19):2354-60.

Leber AW, Knez A, von Ziegler F, et al. Quantification of obstructive and nonobstructive coronary lesions by 64-slice computed tomography: a comparative study with quantitative coronary angiography and intravascular ultrasound.[see comment]. *J Am Coll Cardiol* 2005; 46 (1):147-54.

Leschka S, Alkadhi H, Plass A, et al. Accuracy of MSCT coronary angiography with 64-slice technology: first experience.[see comment]. *Eur Heart J* 2005; 26 (15):1482-7.

Lin GA, Dudley RA, Redberg RF. Cardiologists' use of percutaneous coronary interventions for stable coronary artery disease. *Arch Intern Med* 2007;167(15):1604-9.

MacMahon H, Austin JH, Gamsu G et al. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. *Radiology* 2005; 237 (2):395-400.

Matt D, Scheffel H, Leschka S, et al. Dual-source CT coronary angiography: image quality, mean heart rate, and heart rate variability. *AJR Am J Roentgenol* 2007; 189 (3):567-73.

Mayo Foundation for Medical Education and Research. Coronary Artery Disease. http://www.mayoclinic.com/health/coronary-artery-disease/DS00064 (accessed 8/19/08).

Medical Services Advisory Committee, Australian Government. Computed Tomography Coronary Angiogram. http://www.msac.gov.au/internet/msac/publishing.nsf/Content/app1105-1. 2007. (accessed Aug 6, 2008).

Medicare Coverage Advisory Committee. Non-invasive diagnostic imaging compared with coronary artery angiography in the diagnosis of coronary artery disease - question scoring sheet. http://www.cms.hhs.gov/mcd/viewmcac.asp?from2=viewmcac.asp&where=index&mid=34& 2006. (accessed 2008).

Meijboom WB, Mollet NR, Van Mieghem CA, et al. [3] 64-Slice CT coronary angiography in patients with non-ST elevation acute coronary syndrome.[see comment]. *Heart* 2007; 93 (11):1386-92.

Meijboom WB, van Mieghem CA, Mollet NR, et al. [1] 64-slice computed tomography coronary angiography in patients with high, intermediate, or low pretest probability of significant coronary artery disease. *J Am Coll Cardiol* 2007; 50 (15):1469-75.

Meijboom WB, Weustink AC, Pugliese F, et al. [2] Comparison of diagnostic accuracy of 64-slice computed tomography coronary angiography in women versus men with angina pectoris. *Am J Cardiol* 2007; 100 (10):1532-7.

Mettler FA, Jr., Huda W, Yoshizumi TT, et al. Effective doses in radiology and diagnostic nuclear medicine: a catalog. *Radiology* 2008; 248 (1):254-63.

Miller DD, Shaw LJ. Coronary artery disease: diagnostic and prognostic models for reducing patient risk. *J Cardiovasc Nurs* 2006; 21 (6 Suppl 1):S2-16; quiz S7-9.

Miller JM, Rochitte CE, Dewey M et al. Diagnostic performance of coronary angiography by 64-row CT. *N Engl J Med* 2008; 359 (22):2324-36.

Min JK, Shaw LJ, Berman DS, et al. [1] Costs and clinical outcomes in individuals without known coronary artery disease undergoing coronary computed tomographic angiography from an analysis of Medicare category III transaction codes. *Am J Cardiol* 2008; 102 (6):672-8.

Min JK, Kang N, Shaw LJ, et al. [2] Costs and clinical outcomes after coronary multidetector CT angiography in patients without known coronary artery disease: comparison to myocardial perfusion SPECT. *Radiology* 2008; 249 (1):62-70.

Mollet NR, Cademartiri F, van Mieghem CA, et al. High-resolution spiral computed tomography coronary angiography in patients referred for diagnostic conventional coronary angiography.[see comment]. *Circulation* 2005; 112 (15):2318-23.

Mowatt G, Cummins E, Waugh N, et al. Systematic review of the clinical effectiveness and cost-effectiveness of 64-slice or higher computed tomography angiography as an alternative to invasive coronary angiography in the investigation of coronary artery disease. *Health Technol Assess* 2008; 12 (17):iii-iv, ix-143.

Mowatt G, Vale L, Brazzelli M, et al. Systematic review of the effectiveness and cost-effectiveness, and economic evaluation, of myocardial perfusion scintigraphy for the diagnosis and management of angina and myocardial infarction. *Health Technol Assess* 2004; 8 (30):iii-iv, 1-207.

Muhlenbruch G, Seyfarth T, Soo CS, et al. Diagnostic value of 64-slice multi-detector row cardiac CTA in symptomatic patients. *Eur Radiol* 2007; 17 (3):603-9.

National Institute for Health and Clinical Excellence. 64 slice CT coronary angiography for angina and suspected coronary artery disease. http://www.nice.org.uk/guidance/index.jsp?action=byID&o=11953 . (accessed 9/2/08).

Nikolaou K, Knez A, Rist C, et al. Accuracy of 64-MDCT in the diagnosis of ischemic heart disease. *AJR Am J Roentgenol* 2006; 187 (1):111-7.

Ontario Health Technology Advisory Committee. Multi-Detector Computed Tomography Angiography for Coronary Artery Disease. http://www.health.gov.on.ca/english/providers/program/ohtac/tech/reviews/sum_mdct_20070926.html. 2005. (accessed Aug 6, 2008).

Oncel D, Oncel G, Tastan A, et al. Detection of significant coronary artery stenosis with 64-section MDCT angiography. *Eur J Radiol* 2007; 62 (3):394-405.

Ong TK, Chin SP, Liew CK, et al. Accuracy of 64-row multidetector computed tomography in detecting coronary artery disease in 134 symptomatic patients: influence of calcification. *Am Heart J* 2006; 151 (6):1323 e1-6.

Onuma Y, Tanabe K, Nakazawa G, et al. Noncardiac findings in cardiac imaging with multidetector computed tomography. *J Am Coll Cardiol* 2006; 48 (2):402-6.

Patterson RE, Eisner RL, Horowitz SF. Comparison of cost-effectiveness and utility of exercise ECG, single photon emission computed tomography, positron emission tomography, and coronary angiography for diagnosis of coronary artery disease. *Circulation* 1995; 91:54-65.

Plass A, Grunenfelder J, Leschka S et al. Coronary artery imaging with 64-slice computed tomography from cardiac surgical perspective.[see comment]. *Eur J Cardiothorac Surg* 2006; 30 (1):109-16.

Pugliese F, Mollet NR, Hunink MG, et al. Diagnostic performance of coronary CT angiography by using different generations of multisection scanners: single-center experience. *Radiology* 2008; 246 (2):384-93.

Pugliese F, Mollet NR, Runza G, et al. Diagnostic accuracy of non-invasive 64-slice CT coronary angiography in patients with stable angina pectoris. *Eur Radiol* 2006; 16 (3):575-82.

Pundziute G, Schuijf JD, Jukema JW, et al. Prognostic value of multislice computed tomography coronary angiography in patients with known or suspected coronary artery disease.[see comment]. *J Am Coll Cardiol* 2007; 49 (1):62-70.

Pundziute G, Schuijf JD, Jukema JW, et al. Gender influence on the diagnostic accuracy of 64-slice multislice computed tomography coronary angiography for detection of obstructive coronary artery disease. *Heart* 2008; 94 (1):48-52.

Raff GL, Gallagher MJ, O'Neill WW, et al. Diagnostic accuracy of noninvasive coronary angiography using 64-slice spiral computed tomography.[see comment]. *J Am Coll Cardiol* 2005; 46 (3):552-7.

Ratib O. Coronary CT angiography: overuse, underuse or adequate use? *Kardiovaskuläre Medizin* 2008; 11(6):202–208.

Reynolds MR, Neil N, Ho KK, et al. Clinical and economic outcomes of multivessel coronary stenting compared with bypass surgery: a single-center US experience. *Am Heart J* 2003;145(2):334-42.

Ropers D, Rixe J, Anders K, et al. Usefulness of multidetector row spiral computed tomography with 64- x 0.6-mm collimation and 330-ms rotation for the noninvasive detection of significant coronary artery stenoses. *Am J Cardiol* 2006; 97 (3):343-8.

Ropers U, Ropers D, Pflederer T, et al. Influence of heart rate on the diagnostic accuracy of dual-source computed tomography coronary angiography. *J Am Coll Cardiol* 2007; 50 (25):2393-8.

Rubinshtein R, Halon DA, Gaspar T, et al. [3] Impact of 64-slice cardiac computed tomographic angiography on clinical decision-making in emergency department patients with chest pain of possible myocardial ischemic origin. *Am J Cardiol* 2007; 100 (10):1522-6.

Rubinshtein R, Halon DA, Gaspar T, et al. [2] Usefulness of 64-slice cardiac computed tomographic angiography for diagnosing acute coronary syndromes and predicting clinical outcome in emergency department patients with chest pain of uncertain origin.[see comment]. *Circulation* 2007; 115 (13):1762-8.

Rubinshtein R, Halon DA, Gaspar T, et al. [1] Usefulness of 64-slice multidetector computed tomography in diagnostic triage of patients with chest pain and negative or nondiagnostic exercise treadmill test result. *Am J Cardiol* 2007; 99 (7):925-9.

Sallach SM, Nowak R, Hudson MP, et al. A change in serum myoglobin to detect acute myocardial infarction in patients with normal troponin I levels. *Am J Cardiol* 2004;94(7):864-7.

Sato A, Hiroe M, Tamura M, et al. Quantitative measures of coronary stenosis severity by 64-Slice CT angiography and relation to physiologic significance of perfusion in nonobese patients: comparison with stress myocardial perfusion imaging. *J Nucl Med* 2008; 49 (4):564-72.

Savino G, Herzog C, Costello P, et al. 64 slice cardiovascular CT in the emergency department: concepts and first experiences. *Radiol Med* 2006; 111 (4):481-96.

Scheffel H, Alkadhi H, Plass A, et al. Accuracy of dual-source CT coronary angiography: First experience in a high pre-test probability population without heart rate control. *Eur Radiol* 2006; 16 (12):2739-47.

Schlosser T, Mohrs OK, Magedanz A, et al. Noninvasive coronary angiography using 64-detector-row computed tomography in patients with a low to moderate pretest probability of significant coronary artery disease. *Acta Radiol* 2007; 48 (3):300-7.

Schroeder S, Achenbch S, Bengel F, et al. Cardiac computed tomography: indications, applications, limitations, and training requirements. *Eur Heart J* 2008; (29):531–556.

Schuijf JD, Pundziute G, Jukema JW, et al. [1] Diagnostic accuracy of 64-slice multislice computed tomography in the noninvasive evaluation of significant coronary artery disease. *Am J Cardiol* 2006; 98 (2):145-8.

Schuijf JD, Wijns W, Jukema JW, et al. [3] Relationship between noninvasive coronary angiography with multi-slice computed tomography and myocardial perfusion imaging.[see comment]. *J Am Coll Cardiol* 2006; 48 (12):2508-14.

Schuijf JD, Wijns W, Jukema JW, et al. [2] A comparative regional analysis of coronary atherosclerosis and calcium score on multislice CT versus myocardial perfusion on SPECT. *J Nucl Med* 2006; 47 (11):1749-55.

Shabestari AA, Abdi S, Akhlaghpoor S, et al. Diagnostic performance of 64-channel multislice computed tomography in assessment of significant coronary artery disease in symptomatic subjects. *Am J Cardiol* 2007; 99 (12):1656-61.

Shapiro MD, Butler J, Rieber J, et al. Analytic approaches to establish the diagnostic accuracy of coronary computed tomography angiography as a tool for clinical decision making.[see comment]. *Am J Cardiol* 2007; 99 (8):1122-7.

Shuman WP, Branch KR, May JM, et al. Prospective versus retrospective ECG gating for 64-detector CT of the coronary arteries: comparison of image quality and patient radiation dose. *Radiology* 2008; 248 (2):431-7.

Song F, Khan KS, Dinnes J et al. Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. *Int J Epidemiol* 2002; 31 (1):88-95.

Stolzmann P, Scheffel H, Leschka S et al. Influence of calcifications on diagnostic accuracy of coronary CT angiography using prospective ECG triggering. *AJR Am J Roentgenol* 2008; 191 (6):1684-9.

Topol EJ, Nissen SE. Our preoccupation with coronary luminology: the disassociation between clinical and angiographic findings in ischemic heart disease. *Circulation* 1995; 92 (8):2333-42.

Valenza T. "CCTA: Way out of the view box." *Imaging Economics* 2006. Ascend Media, LLC: Overland Park, KS. http://www.imagingeconomics.com/issues/articles/2006-12_01.asp (accessed October 31, 2008).

Van Gelder, RE, Nio CY, Florie J, et al. Computed tomographic colonography compared with colonoscopy in patients at increased risk for colorectal cancer. *Gastroenterology* 2004; 127 (1):41-8.

Ward RP, Goonewardena SN, Lammertin G, et al. Comparison of the frequency of abnormal cardiac findings by echocardiography in patients with and without peripheral arterial disease. *Am J Cardiol* 2007;99(4):499-503.

Weisbord SD. Iodinated contrast media and the kidney. *Rev Cardiovasc Med* 2008; 9 Suppl 1:S14-23.

Whiting P, Rutjes AW, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; 3:25.

Zalenski RJ, McCarren M, Roberts R, et al. An evaluation of a chest pain diagnostic protocol to exclude acute cardiac ischemia in the emergency department. *Arch Intern Med* 1997;157(10):1085-91.

Zamora J, Abraira V, Muriel A, et al. Meta-DiSc: a software for meta-analysis of test accuracy data. In: BMC Medical Research Methodology 6:31 end., Series Meta-DiSc: a software for meta-analysis of test accuracy data. 2006.

SYSTEMATIC REVIEW TABLES

Table 1. Characteristics of excluded studies.

Author	Year	Sample Size	Referent Standard	Mean Age	% Male	Reason for Exclusion
Cademartiri [4]	2008	170	ICA	58	73	No per-patient analysis
Gaemperli	2007	100	SPECT	61	70	For detection of functional abnormalities only
Meijboom	2007	254	ICA	56	51	Overlap with another study sample
Mollet	2007	62	ICA	59	73	Mixture of 16- and 64-slice CCTA
Ong	2006	134	ICA	55	73	No per-patient analysis
Rubinshtein [2]	2007	58	ICA/Other	56	64	Identical to Rubenshtein (3) study
Sato	2008	104	SPECT	67	73	No per patient analysis
Schlosser	2007	63	ICA	62	65	No per patient analysis
Schuijf [2]	2006	140	SPECT	59	60	No per patient analysis

Table 2. Characteristics of included studies.

Author	Year	Sample Size	Referent Standard	Mean Age	% Male	% Known CAD	% Non-Evaluable	Comments
Achenbach (a)	2008	100	ICA	65	50%	0	18%	Single-source CT
Achenbach (b)	2008	100	ICA	61	64%	0	3%	Dual-source CT
Bayrak	2008	100	ICA	58	70%	16%	0	
Cademartiri [1]	2008	145	ICA	63	63%	Unk	8%	
Cademartiri [3]	2007	72	ICA	54	53%	Unk	0	
Ehara	2006	69	ICA	67	75%	57%	3%	
Fine	2006	66	ICA	62	48%	5%	6%	
Gallagher	2007	85	Clin Dx	49	53%	0%	8%	
Ghostine	2006	66	ICA	69	61%	0	0	
Goldstein (a)	2007	99	ICA/ Other	48	42%	0	0	CCTA arm
Goldstein (b)	2007	98	ICA/ Other	51	57%	0	0	SOC arm
Hacker	2007	38	SPECT/ ICA	62	74%	32%	0	Functionally relevant stenosis
Hoffmann	2006	103	Clin Dx	54	60%	10%	0	
Hollander	2007	54	Unk	47	46%	2%	0	
Husmann (a)	2008	34	ICA	63	29%	Unk	0	Low risk
Husmann (b)	2008	29	ICA	64	66%	Unk	0	Intermed risk
Johnson [1]	2007	55	ICA/ Other	67	64%	Unk	2%	
Johnson [2]	2007	35	ICA	60	69%	40%	0	
Leber	2007	90	ICA	58	63%	0	2%	
Leber	2005	59	ICA	64	--	17%	7%	
Leschka	2005	67	ICA	60	75%	Unk	0	
Meijboom (2a)	2007	123	ICA	62	0%	12%	0	Women
Meijboom (2b)	2007	279	ICA	58	100%	10%	0	Men
Meijboom (3a)	2007	33	ICA	58	70%	3%	0	low risk
Meijboom (3b)	2007	71	ICA	59	73%	24%	0	high risk
Mollet	2005	52	ICA	60	65%	Unk	0	
Muhlenbruch	2007	51	ICA	59	76%	Unk	0	
Nikolaou	2006	72	ICA	64	82%	40%	6%	
Oncel	2007	80	ICA	56	76%	0	0	
Plass	2006	50	ICA	66	78%	Unk	0	
Pugliese	2006	35	ICA	61	60%	9%	0	
Pugliese	2008	51	ICA	59	76%	Unk	0	
Pundziute	2008				--	--	--	
Pundziute	2008	103	ICA	60	49%	33%	3%	
Raff	2005	70	ICA	59	76%	Unk	0	
Ropers D	2006	84	ICA	58	62%	Unk	4%	

Table 2. Characteristics of included studies.

Author	Year	Sample Size	Referent Standard	Mean Age	% Male	% Known CAD	% Non-Evaluable	Comments
Ropers U	2007	100	ICA	61	63%	0	3%	
Rubinshtein (1)	2007	100	ICA	56	57%	0	3%	
Rubinshtein (3)	2007	58	ICA/ Other	56	64%	38%	0	Combination ICA/Dx protocol
Savino	2006	23	ICA	56	61%	0	0	
Scheffel	2006	30	ICA	63	80%	Unk	0	
Schuijf (1)	2006	60	ICA	60	77%	55%	2%	
Schuijf (3)	2006	58	ICA	63	66%	0	0	
Shabestiri	2007	143	ICA	63	72%	Unk	3%	
Shapiro	2007	37	ICA	63	78%	32%	14%	

Table 3. Studies examining prognostic ability of 64-slice or better CCTA based on clinical follow-up.

Author	Year	Study Type	Setting	Sample Size	Age (Mean, SD)	% Male	CAD Risk	Follow -Up	Diagnosis Method	Major Findings
Danciu	2007	Case series	OP	421	64 (12)	63%	Intermediate	12 months	SPECT, ICA, MACE	>80% medically managed; event rate 0.3% in low risk group vs. 70.5% for ICA
Goldstein	2007	RCT	ED	99	48 (11)	43%	Very low	6 months	ICA, repeat testing (MACE)	CCTA correctly and definitively diagnosed 94 of 99 (95%)
Gallagher	2007	Clinical practice algorithm	ED	85	49 (11)	53%	Low	30 days	ICA, record review, clinical follow-up	No events recorded; CCTA positive in 6 of 7 confirmed cases of ACS
Hoffmann	2006	Validation cohort	ED	103	54 (12)	60%	Low	Mean: 5.2 months	Record review (index visit only, ACS)	Sensitivity 100% for ACS, specificity 82%
Hollander	2007	Clinical practice algorithm	ED	54	46.5 (8.5)	46%	Low	30 days	Survey, record review (cardiac death/acute MI)	No events recorded; CAD confirmed in 4 of 6 CCTA-positive patients
Johnson [1]	2007	Clinical practice algorithm	ED	55	67 (10)	64%	N/A	≥5 months	Record review, repeat enzymes	CCTA correctly and definitively diagnosed 51 of 55 (93%)
Pundziute	2007	Clinical practice algorithm	OP	100	59 (12)	73%	Intermediate	Mean: 16 months	Record review, clinic visits, survey (MACE)	1-yr event rate 0% in CCTA (-) patients; 30% in CCTA (+)
Rubinshtein [3]	2007	Clinical practice algorithm	ED	58	56 (10)	64%	Intermediate	12 months	Altered strategies, f/u survey	Revised ACS diagnosis, canceled hospitalization in ~45%; no events in CCTA (-)
Savino	2006	Validation cohort	ED	23	56 (13)	61%	N/A	ED visit only	Record review	All moderate/severe stenoses on CCTA confirmed by ICA

CAD: coronary artery disease; RCT: randomized controlled trial; MACE: major adverse cardiovascular event; CCTA: coronary computed tomographic angiography

Table 4. Sensitivity and specificity (intent-to-diagnose analysis).

Author	Year	TP	FP	TN	FN	Sensitivity	(95% CI)	Specificity	(95% CI)	PPV	NPV
Achenbach	2008	34	27	35	7	0.83		0.80		0.79	0.83
Achenbach	2008	39	9	51	1	0.98		0.89		0.87	0.98
Bayrak	2008	64	4	32	0	1.00	(.99-1.00)	0.89	(.80-.98)	0.94	1.00
Cademartiri [1]	2008	82	32	29	2	0.98	(.91-.99)	0.58	(.43-.71)	0.80	0.94
Cademartiri [2]	2007	20	1	51	0	1.00	(.83-1.00)	0.98	(.89-.99)	0.95	1.00
Ehara	2006	59	3	6	1	0.98		0.86		0.98	0.86
Fine	2006	35	5	24	2	0.95		0.96		0.97	0.92
Ghostine	2006	28	2	35	1	0.97	(.82-1.00)	0.95	(.82-.99)	0.93	0.97
Hacker	2007	11	7	10	2	0.85		0.59		0.61	0.83
Husmann	2008	9	5	19	1	0.90	(.56-.99)	0.79	(.58-.93)	0.64	0.95
Husmann	2008	14	1	12	2	0.88	(.62-.99)	0.92	(.64-1.00)	0.93	0.86
Johnson	2007	17	2	16	0	1.00	(.83-1.00)	0.89	(.65-.98)	0.89	1.00
Leber	2007	20	9	60	1	0.95	(.76-.99)	0.90	(.80-.95)	0.74	0.98
Leber	2005	22	7	17	3	0.88		0.85		0.88	0.85
Leschka	2005	47	0	20	0	1.00		1.00		1.00	1.00
Meijboom	2007	63	15	45	0	1.00	(.93-1.00)	0.75	(.62-.85)	0.81	1.00
Meijboom	2007	188	9	80	2	0.99	(.96-1.00)	0.90	(.81-.95)	0.95	0.98
Meijboom	2007	28	1	4	0	1.00	(.85-1.00)	0.80	(.30-.99)	0.93	1.00
Meijboom	2007	60	3	8	0	1.00	(.93-1.00)	0.73	(.39-.93)	0.95	1.00
Mollet	2005	38	2	12	0	1.00	(.91-1.00)	0.92	(.67-.99)	0.97	1.00
Muhlenbruch	2007	44	3	3	1	0.98	(.88-.99)	0.50	(.11-.88)	0.94	0.75
Nikolaou	2006	38	10	23	1	0.97		0.79		0.86	0.96
Oncel	2007	62	0	18	0	1.00		1.00		1.00	1.00
Plass	2006	40	1	9	0	1.00		0.90		0.98	1.00
Pugliese	2006	25	1	9	0	1.00	(.87-1.00)	0.90	(.59-.98)	0.96	1.00
Pugliese	2008	38	0	13	0	1.00	(.88-1.00)	1.00	(.71-1.00)	1.00	1.00
Pundziute	2008	53	7	42	1	0.98	(.95-1.00)	0.91	(.83-.99)	0.93	0.98
Raff	2005	38	3	27	2	0.95		0.90		0.93	0.93
Ropers D	2006	25	8	50	1	0.96	(.80-1.00)	0.91	(.80-.97)	0.83	0.98
Ropers U	2007	41	11	47	1	0.98	(.88-1.00)	0.81	(.69-.89)	0.79	0.98
Rubinshtein [1]	2007	26	6	70	1	0.96		0.96		0.90	0.99
Scheffel	2006	14	0	15	1	0.93	(.68-1.00)	1.00	(.78-1.00)	1.00	0.94
Schuijf [1]	2006	29	2	28	2	0.94	(.86-1.00)	0.97	(.91-1.00)	0.97	0.93
Schuijf [3]	2006	27	6	25	0	1.00		0.81		0.82	1.00
Shabestiri	2007	104	15	20	4	0.96	(.91-.99)	0.67	(.47-.83)	0.91	0.83
Shapiro	2007	28	3	5	1	0.97	(.80-1.00)	.63	(.20-.93)	.90	0.83

Table 5. Reports of incidental findings on multi-slice CCTA.

Author	Year	Sample Size	Mean Age	% Males	% with Incidental Findings	% with Significant Findings	% with Therapeutic Consequences
Cademartiri [4]	2007	670	60	57	79	12	2
Dewey	2007	108	63	74	15	5	1
Gil	2007	258	54	78	--	56	--
Kirsch	2007	100	63	68	67	11	--
Onuma	2006	503	66	76	58	23	4

NOTE: "Therapeutic consequences" relate to findings that triggered treatment and/or resolution.

APPENDICES

APPENDIX A:

CLINICAL GUIDELINES

(in separate document, available upon request)

APPENDIX B:

LITERATURE SEARCH STRATEGY

The search strategy for MEDLINE was:

1. coronary artery disease [MeSH Terms]
2. coronary stenosis [MeSH Terms]
3. coronary disease [MeSH Terms]
4. 1 OR 2 OR 3
5. coronary angiography [MeSH Terms]
6. tomography, x-ray computed [MeSH Terms]
7. tomography, spiral computed [MeSH Terms]
8. 64-slice [keyword]
9. 5 OR 6 OR 7 OR 8
10. sensitivity and specificity[MeSH Terms]
11. predictive value of tests[MeSH Terms]
12. prospective studies[MeSH Terms]
13. 10 OR 11 OR 12
14. 4 AND 9 AND 13

The search strategy for EMBASE was:
1. coronary artery disease
2. coronary stenosis
3. 1 OR 2
4. angiography
5. computed tomography
6. 4 OR 5
7. sensitivity
8. predictive
9. 7 OR 8
10.[2005-2008]/py
11. 3 AND 6 AND 9 AND 10

The Cochrane Library was searched using the terms "angiography", "coronary angiography", or "computed tomography angiography"

APPENDIX C

MODIFIED QUADAS TOOL & ASSESSMENT OF STUDY QUALITY

ICER Appraisal of Coronary CT Angiography

Modified QUADAS* Quality Checklist
Studies of Diagnostic Accuracy (64-Slice or Higher)

Study ID: Paper # (if multiple):

Assessor Initials: Assessment Date:

Item	Yes	No	Unclear
Mandatory quality items			
1. Was the spectrum of patients representative of the patients who will receive the test in clinical practice?			
2. Were the selection criteria clearly described?			
3. Is the referent standard likely to correctly classify the target condition(s)?			
4. Is the time period between the referent standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?[a]			
5. Did the entire sample or random selection of the sample receive verification of diagnosis with the referent standard?			
6. Did patients receive the same referent standard regardless of the index test result?			
7. Was the referent standard independent of the index test?			
8. Were the index test results interpreted without knowledge of the results of the referent standard?			
9. Were the referent standard results interpreted without knowledge of the results of the index test?			
10. Were the same clinical data available when index test results were interpreted as would be available when the test is used in practice?			
11. Were uninterpretable/intermediate test results reported?			
12. Were withdrawals from the study explained?			
Additional items			
13. Was an established cut-off point used to define stenosis?[b]			
14. Were data on observer variation reported and within an acceptable range?			
15. Were data presented for appropriate groups of patients?[c]			
16. Was true disease prevalence reported or could it be calculated?[d]			

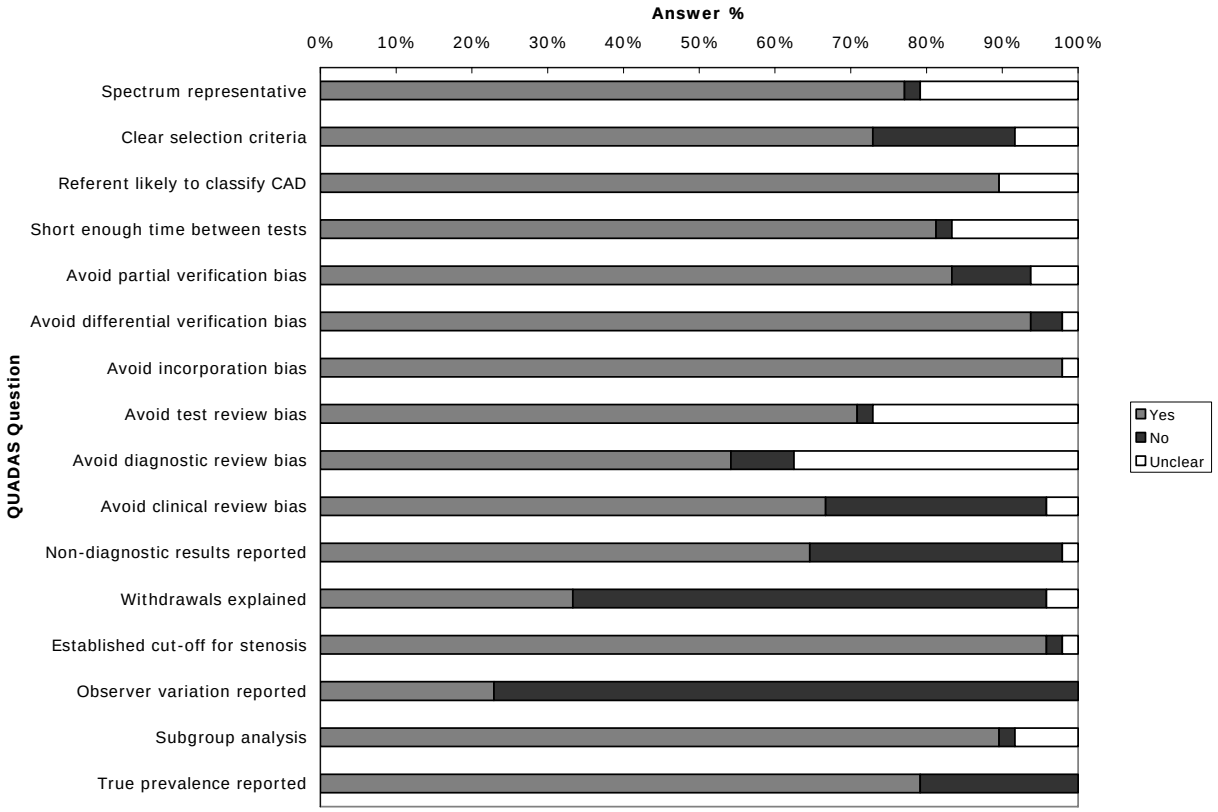
*Whiting P, et al. BMC Medical Research Methodology 2003;3(25):1-13.
NOTE: Original items 8 and 9 were removed from this modified list.
[a]Period of 3 months or less
[b]e.g.,≥50% stenosis
[c]i.e., suspected CAD, low-to-intermediate pretest CAD probability, acute chest pain of unknown origin
[d]Based on number of true-positives on referent standard divided by total sample

Study Quality

A total of 9 studies were rated as "good" quality by the QUADAS tool; in our modification, this represented an answer of "No" or "Unclear" on no more than 3 items. The remainder of studies were rated as "fair", meeting criteria on between 9 and 12 items. As can be seen in Figure 2 below, studies were most often deficient in explaining patient withdrawals and in reporting inter-observer variation; the latter was due in at least some cases to the use of only single blinded reviewers for the both the index and reference tests.

In certain studies, while blinded review of CCTA was clearly described, detail on the methods for review of the ICA results was insufficient or missing entirely. Thirty-five percent of studies did not report the number of patients with non-diagnostic findings. In approximately 30% of studies, the availability of other clinical data was different than in standard practice at the institution, or was unclear. Time between tests, blinding of index reviewers, and independence of the index and reference tests were generally consistently and accurately reported.

Figure 2. Study quality and internal validity, as assessed by modified QUADAS tool.

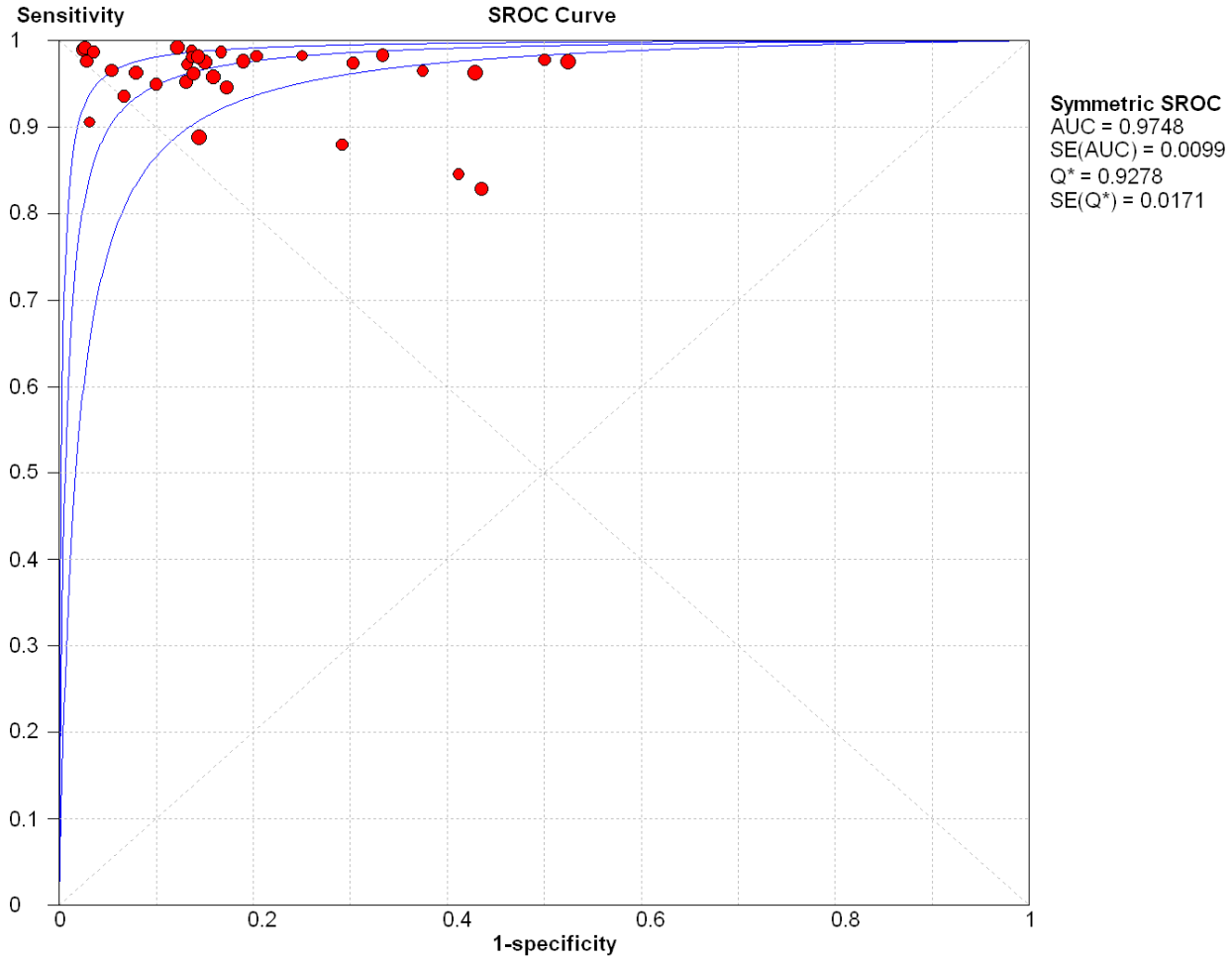


APPENDIX D

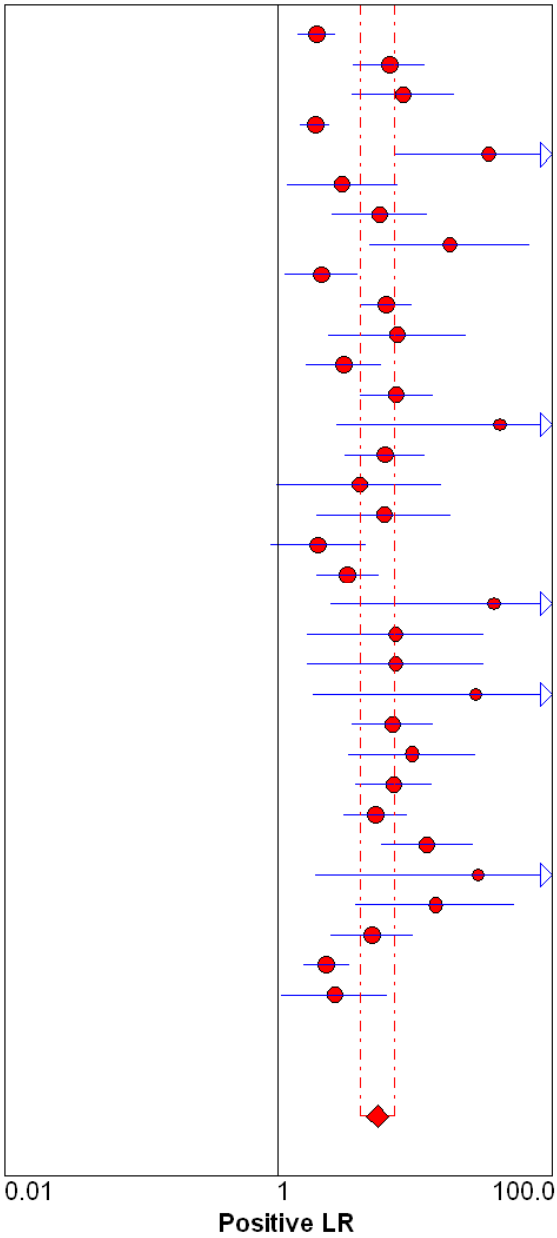
Summary Receiver Operating Characteristic (sROC) Curves & Pooled Likelihood Ratios:

Primary "Intent to Diagnose" Analysis

SROC Curve


Sensitivity
SROC Curve
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
1-specificity
Symmetric SROC
AUC = 0.9748
SE(AUC) = 0.0099
Q* = 0.9278
SE(Q*) = 0.0171

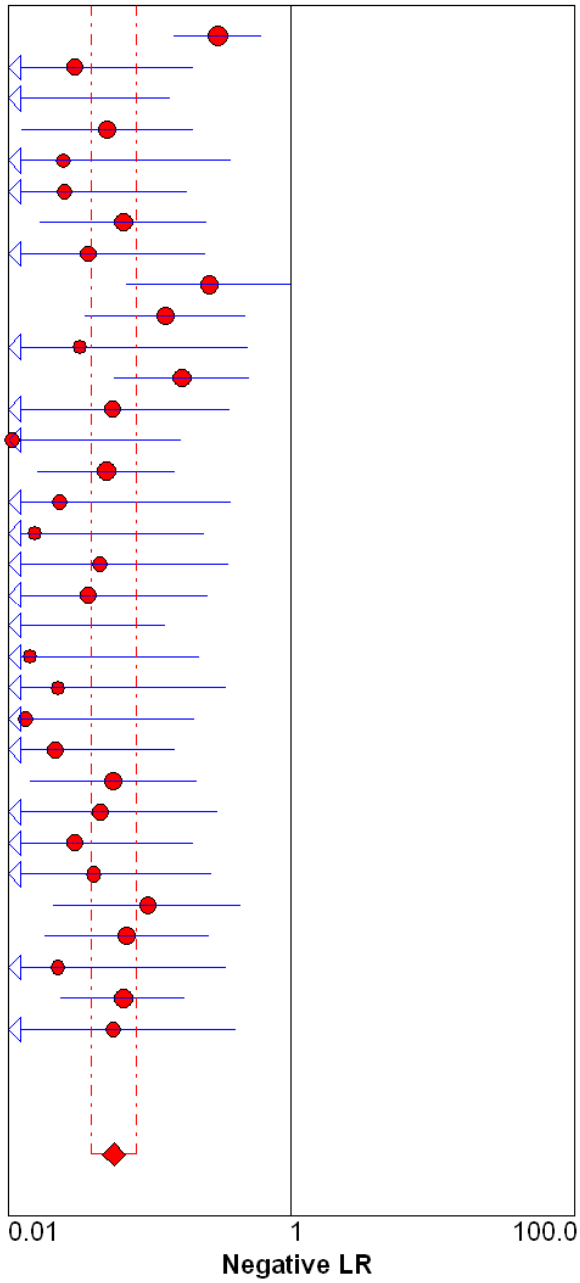
Positive Likelihood Ratio



	Positive LR (95% CI)	
Achenbach (a) 2008	1.90	(1.39 - 2.61)
Achenbach (b) 2008	6.50	(3.55 - 11.90)
Bayrak 2008	8.16	(3.43 - 19.40)
Cademartiri (1) 2008	1.86	(1.46 - 2.37)
Cademartiri (2) 2007	34.49	(7.11 - 167.28)
Ehara 2006	2.95	(1.17 - 7.44)
Fine 2006	5.49	(2.46 - 12.22)
Ghostine 2006	17.86	(4.63 - 68.88)
Hacker 2007	2.05	(1.11 - 3.80)
Husmann 2008	6.18	(3.99 - 9.57)
Johnson (2) 2007	7.39	(2.32 - 23.52)
Leber 2005	3.02	(1.59 - 5.72)
Leber 2007	7.30	(3.94 - 13.53)
Leschka 2005	41.56	(2.69 - 642.89)
Meijboom (2) 2007	6.02	(3.05 - 11.90)
Meijboom (3) 2007	3.93	(0.98 - 15.73)
Mollet 2005	5.92	(1.91 - 18.38)
Muhlenbruch 2007	1.96	(0.88 - 4.36)
Nikolaou 2006	3.22	(1.91 - 5.41)
Oncel 2007	37.70	(2.45 - 581.03)
Plass 2006	7.24	(1.64 - 32.06)
Pugliese 2006	7.19	(1.62 - 31.85)
Pugliese 2008	27.64	(1.82 - 420.48)
Pundziute 2008	6.87	(3.46 - 13.65)
Raff 2005	9.50	(3.24 - 27.86)
Ropers D 2006	6.97	(3.65 - 13.33)
Ropers U 2007	5.15	(3.02 - 8.78)
Rubinshtein (1) 2007	12.20	(5.64 - 26.38)
Scheffel 2006	29.00	(1.89 - 445.86)
Schuijf (1) 2006	14.03	(3.67 - 53.70)
Schuijf (3) 2006	4.84	(2.43 - 9.62)
Shabestiri 2007	2.25	(1.53 - 3.30)
Shapiro 2007	2.57	(1.05 - 6.32)

Random Effects Model
Pooled Positive LR = 5.27 (3.96 to 7.02)
Cochran-Q = 160.81; df = 32 (p = 0.0000)
Inconsistency (I-square) = 80.1 %
Tau-squared = 0.4587

Negative Likelihood Ratio



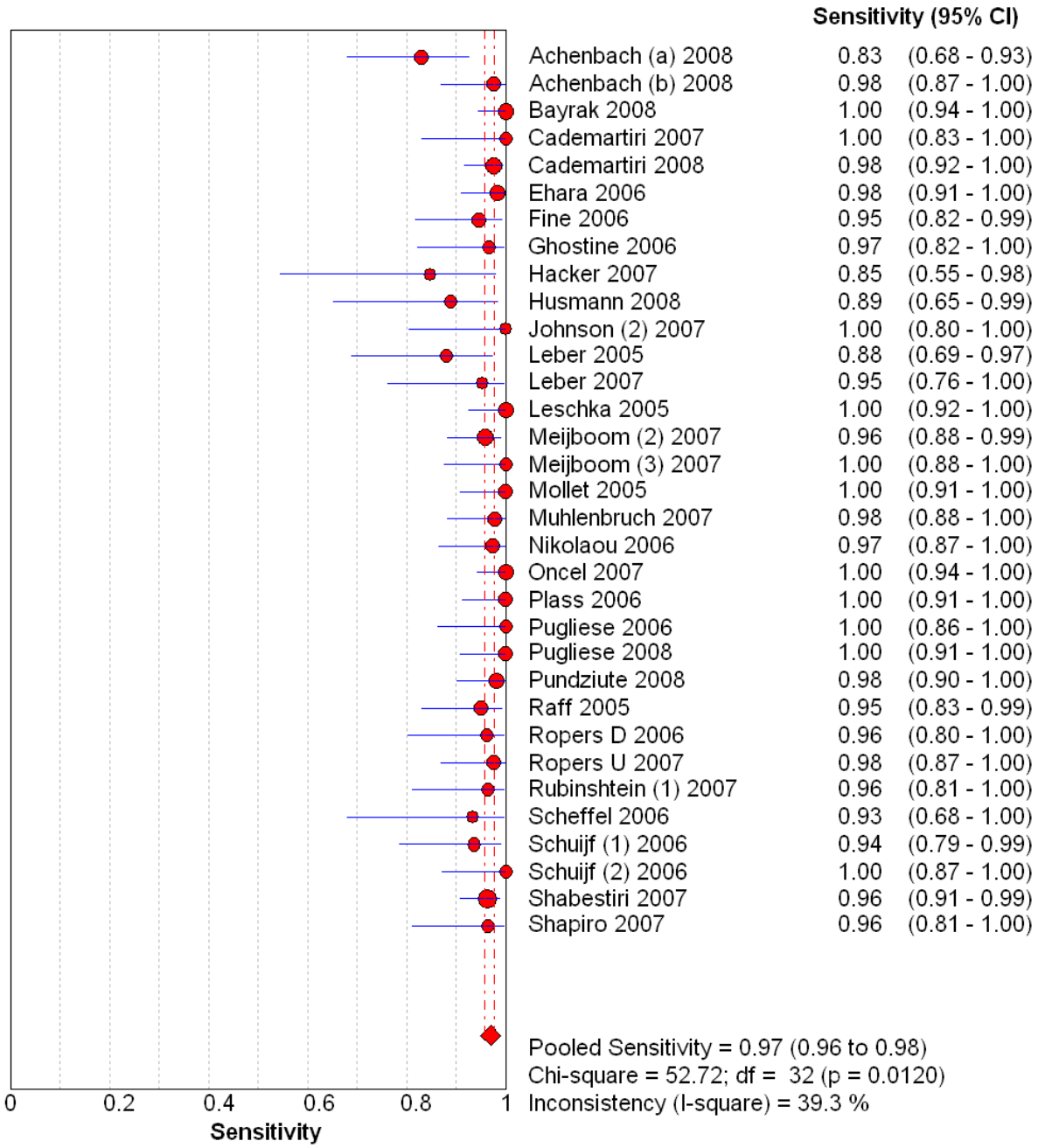
	Negative LR	(95% CI
Achenbach (a) 2008	0.30	(0.15 - 0.61)
Achenbach (b) 2008	0.03	(0.00 - 0.20)
Bayrak 2008	0.01	(0.00 - 0.14)
Cademartiri (1) 2008	0.05	(0.01 - 0.20)
Cademartiri (2) 2007	0.02	(0.00 - 0.38)
Ehara 2006	0.03	(0.00 - 0.18)
Fine 2006	0.07	(0.02 - 0.25)
Ghostine 2006	0.04	(0.01 - 0.25)
Hacker 2007	0.26	(0.07 - 0.99)
Husmann 2008	0.13	(0.04 - 0.48)
Johnson (2) 2007	0.03	(0.00 - 0.49)
Leber 2005	0.17	(0.06 - 0.50)
Leber 2007	0.05	(0.01 - 0.37)
Leschka 2005	0.01	(0.00 - 0.17)
Meijboom (2) 2007	0.05	(0.02 - 0.15)
Meijboom (3) 2007	0.02	(0.00 - 0.37)
Mollet 2005	0.02	(0.00 - 0.24)
Muhlenbruch 2007	0.04	(0.01 - 0.36)
Nikolaou 2006	0.04	(0.01 - 0.26)
Oncel 2007	0.01	(0.00 - 0.13)
Plass 2006	0.01	(0.00 - 0.22)
Pugliese 2006	0.02	(0.00 - 0.35)
Pugliese 2008	0.01	(0.00 - 0.21)
Pundziute 2008	0.02	(0.00 - 0.15)
Raff 2005	0.06	(0.01 - 0.22)
Ropers D 2006	0.04	(0.01 - 0.31)
Ropers U 2007	0.03	(0.00 - 0.20)
Rubinshtein (1) 2007	0.04	(0.01 - 0.28)
Scheffel 2006	0.10	(0.02 - 0.45)
Schuijf (1) 2006	0.07	(0.02 - 0.27)
Schuijf (3) 2006	0.02	(0.00 - 0.35)
Shabestiri 2007	0.06	(0.02 - 0.18)
Shapiro 2007	0.06	(0.01 - 0.41)

Random Effects Model
Pooled Negative LR = 0.06 (0.04 to 0.08)
Cochran-Q = 46.93; df = 32 (p = 0.0431)
Inconsistency (I-square) = 31.8 %
Tau-squared = 0.3254

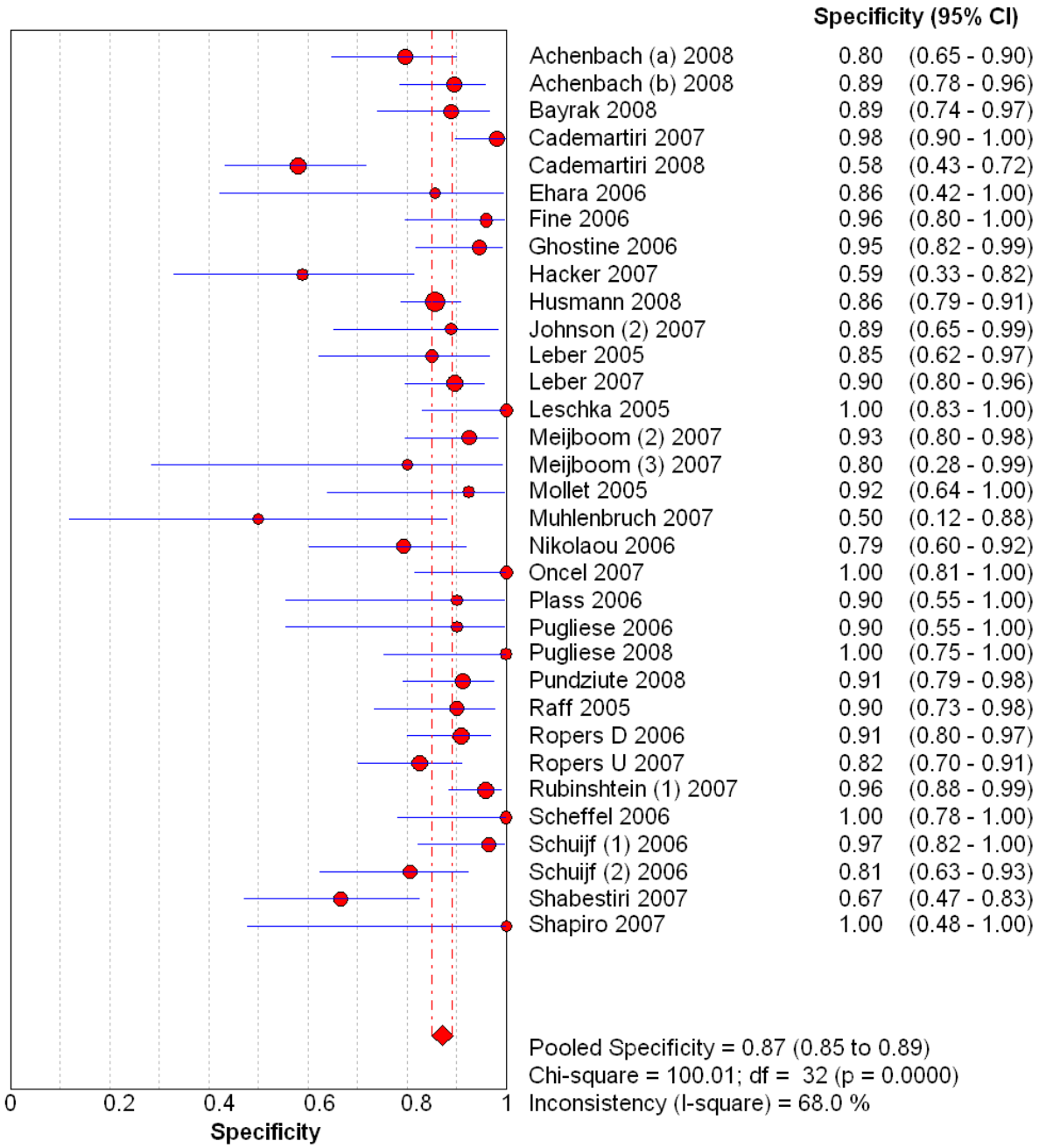
APPENDIX E:

META-ANALYSES OF "AS REPORTED" DATA

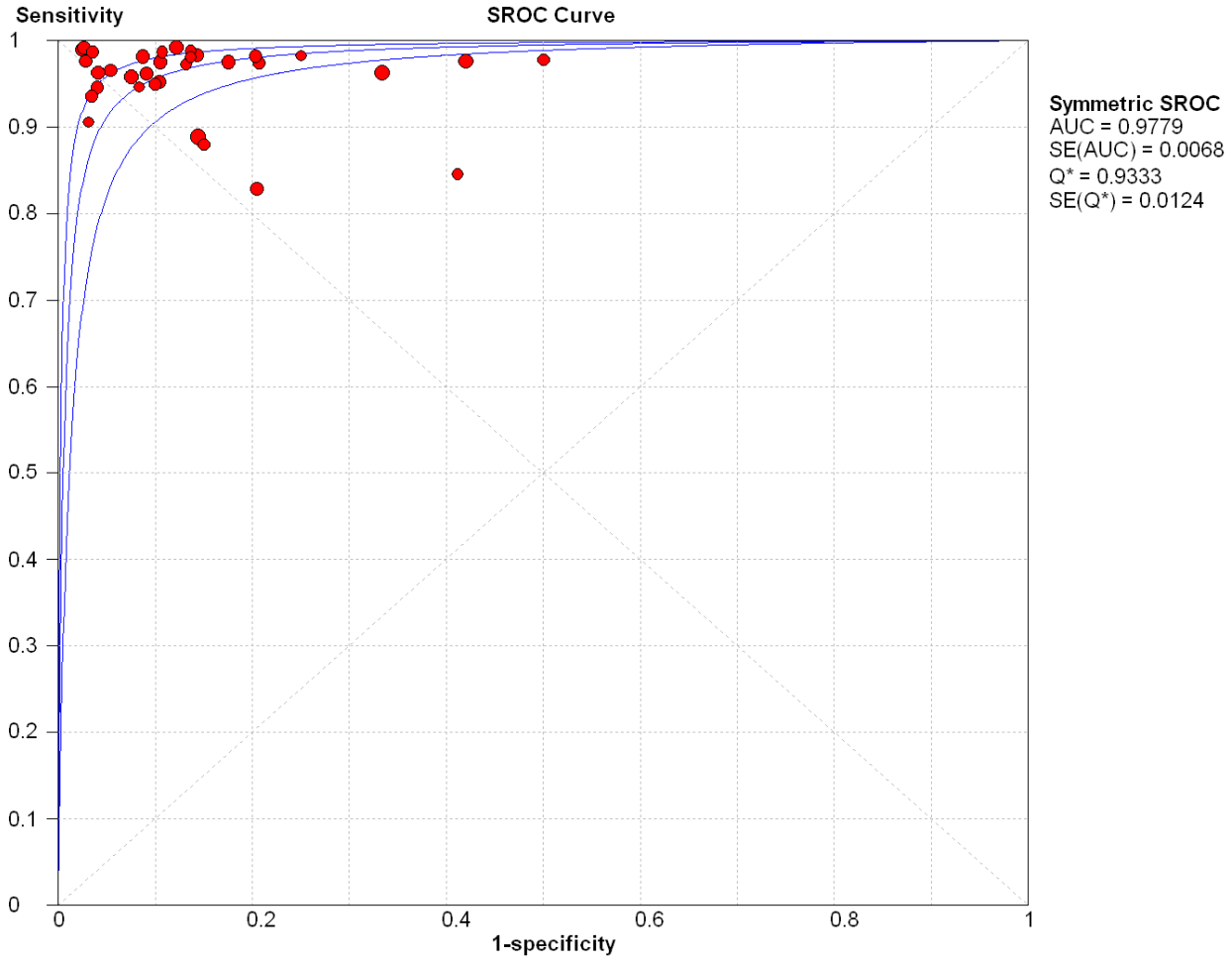
Sensitivity



Specificity



SROC Curve

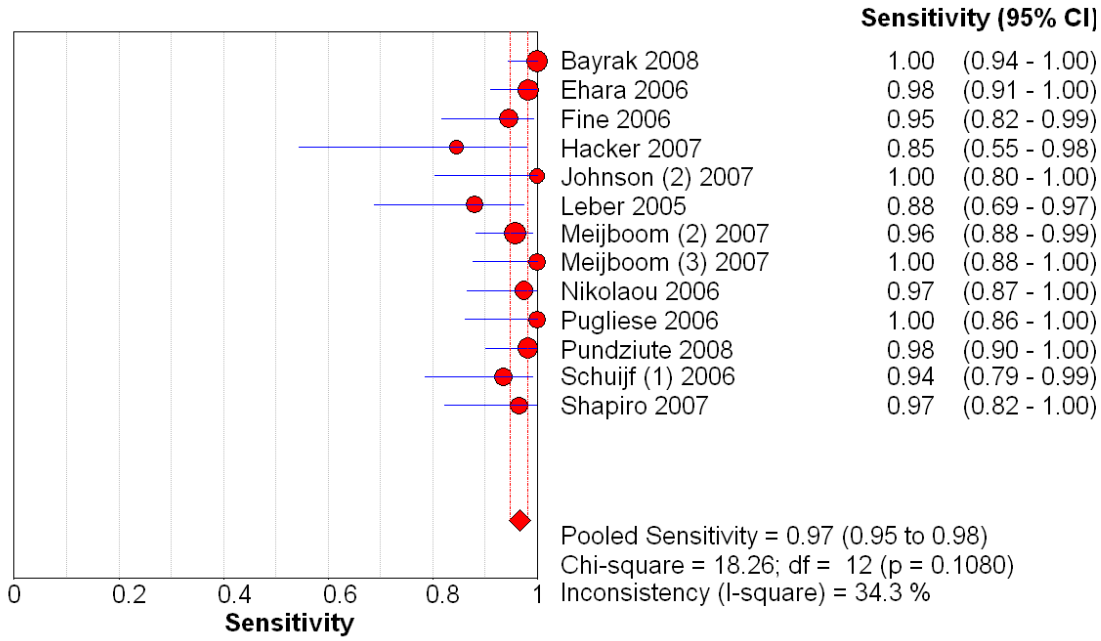

Sensitivity
SROC Curve
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
1-specificity
Symmetric SROC
AUC = 0.9779
SE(AUC) = 0.0068
Q* = 0.9333
SE(Q*) = 0.0124

APPENDIX F:

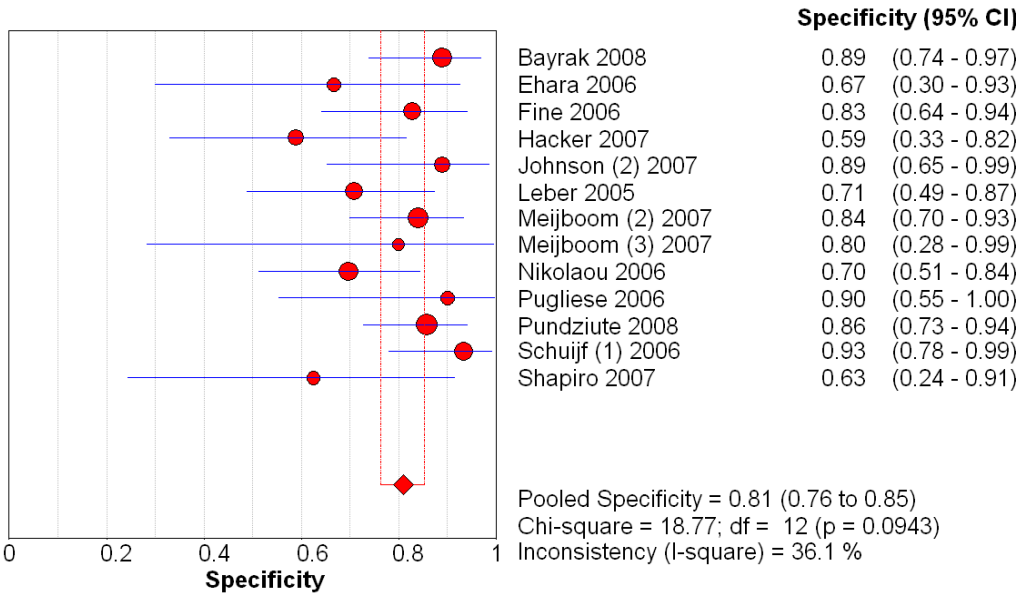
ADDITIONAL ANALYSES BASED ON POPULATIONS WITH AND WITHOUT KNOWN CAD

CAD Known

Sensitivity

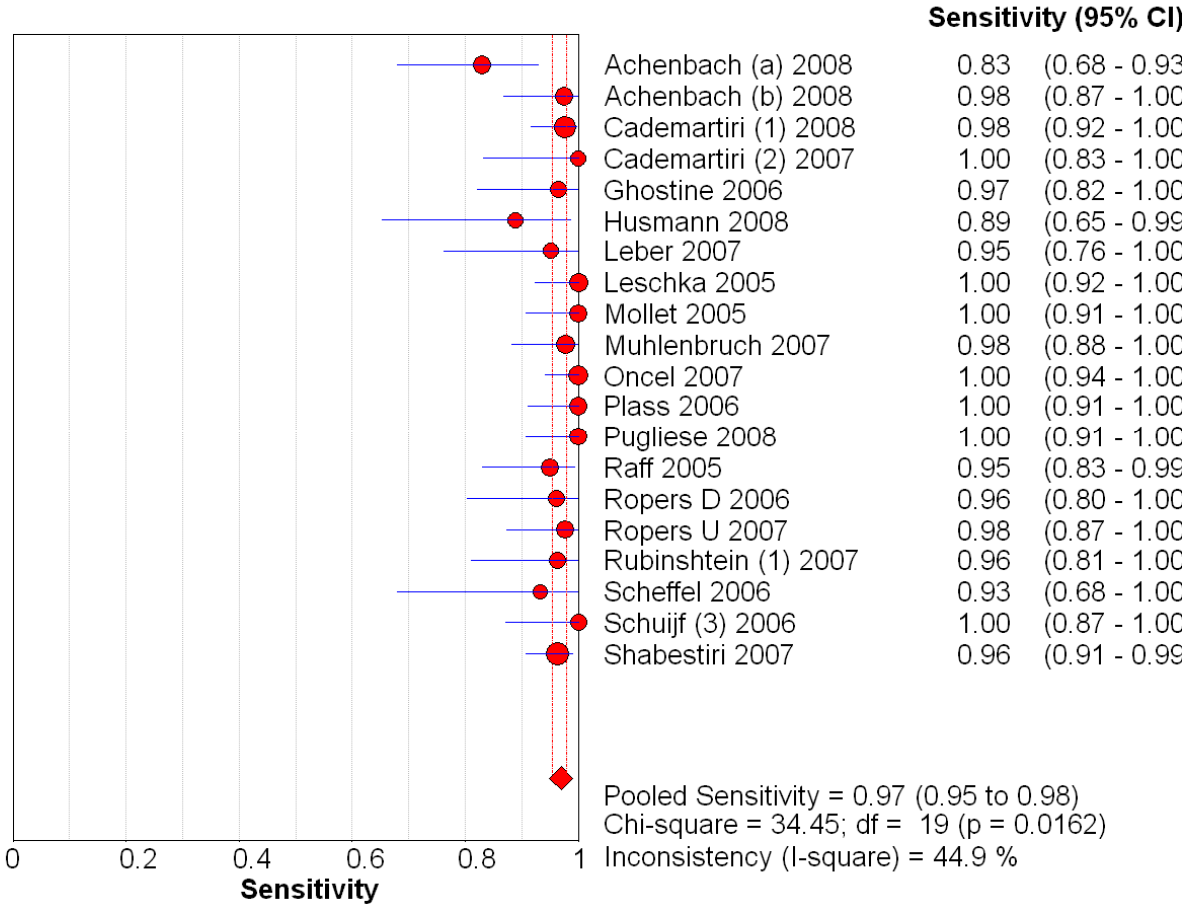


Specificity



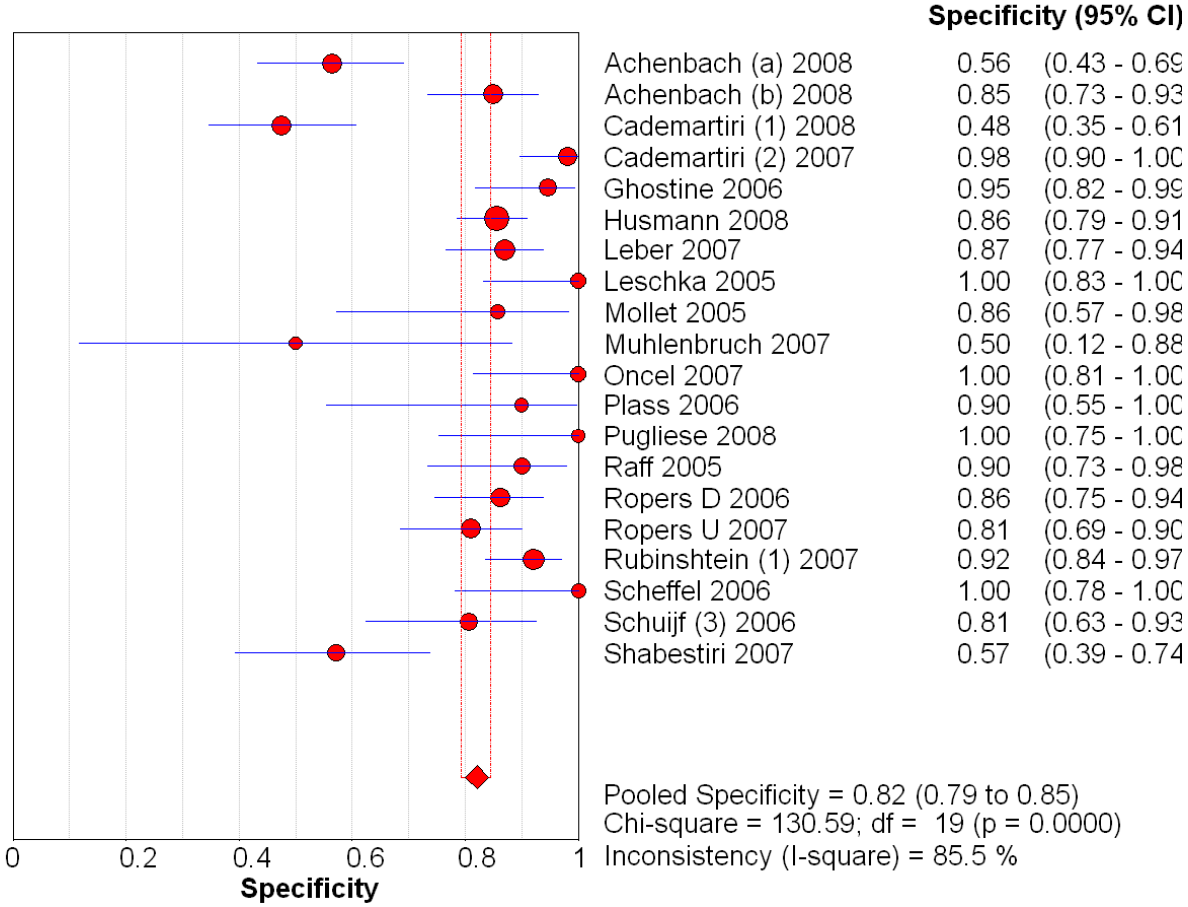
CAD 0 or Unknown

Sensitivity



CAD Unknown

Specificity



APPENDIX G:

ANALYSES OF HETEROGENEITY

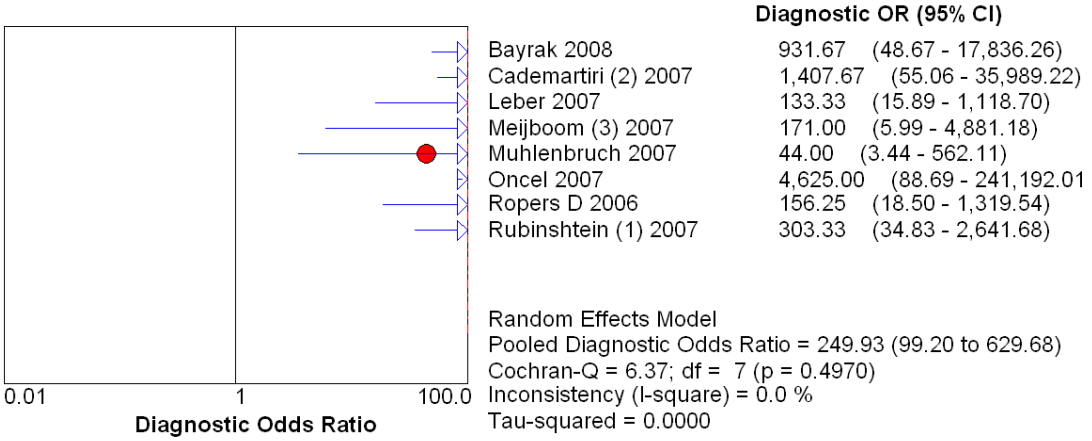
Analysis of Diagnostic Threshold

```
-----------------------------------------------------------------------------
Spearman correlation coefficient: -0.206 p-value= 0.251
(Logit(TPR) vs Logit(FPR)

-----------------------------------------------------------------------------
Moses' model  (D = a + bS)
Weighted regression (Inverse Variance)
 Var      Coeff.         Std. Error        T           p-value
-----------------------------------------------------------------------------
 a         5.106          0.510         10.017         0.0000
 b( 1)    -0.209          0.247          0.846         0.4040

-----------------------------------------------------------------------------
Tau-squared estimate =  1.1394 (Convergence is achieved after 10 iterations)
Restricted Maximum Likelihood estimation (REML)

No. studies =  33
Filter ON (Include = 1 )
Add 1/2 to all cells of the studies with zero
```

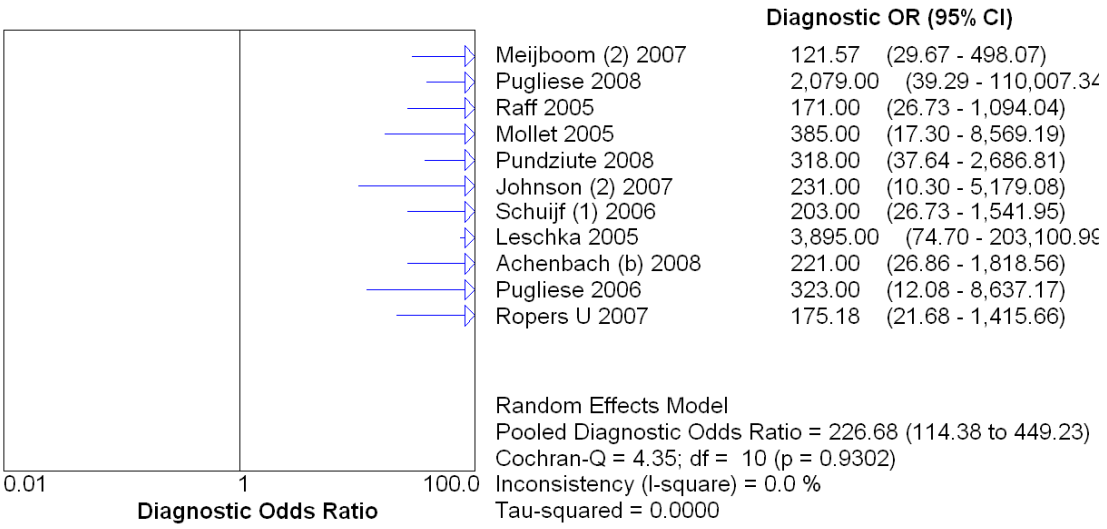
Meta-Regression(Inverse Variance weights)

```
 Var             Coeff.       Std. Err.      p - value          RDOR           [95%CI]
-------------------------------------------------------------------------------------------
Cte.              5.714         2.2043         0.0152           ----            ----
S                -0.164         0.2239         0.4709           ----            ----
CAD               0.138         0.5117         0.7895           1.15         (0.40;3.28)
SS               -0.001         0.0031         0.7370           1.00         (0.99;1.01)
AgeGroup         -1.035         0.3166         0.0029           0.36         (0.19;0.68)
Male              2.267         2.4198         0.3572           9.65         (0.07;1382.56)

-------------------------------------------------------------------------------------------
Tau-squared estimate =  0.6104 (Convergence is achieved after 7 iterations)
Restricted Maximum Likelihood estimation (REML)

No. studies =   33
Filter ON (Include = 1 )
Add 1/2 to all cells of the studies with zero
```

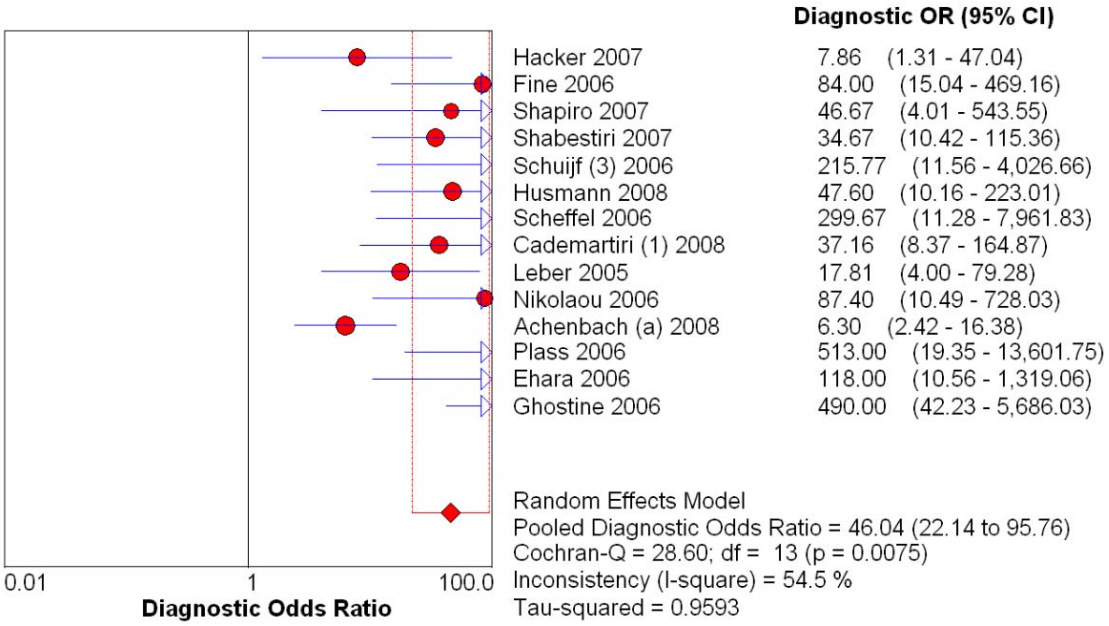
Age Group < 59 years



Age Group 59 to 61 years



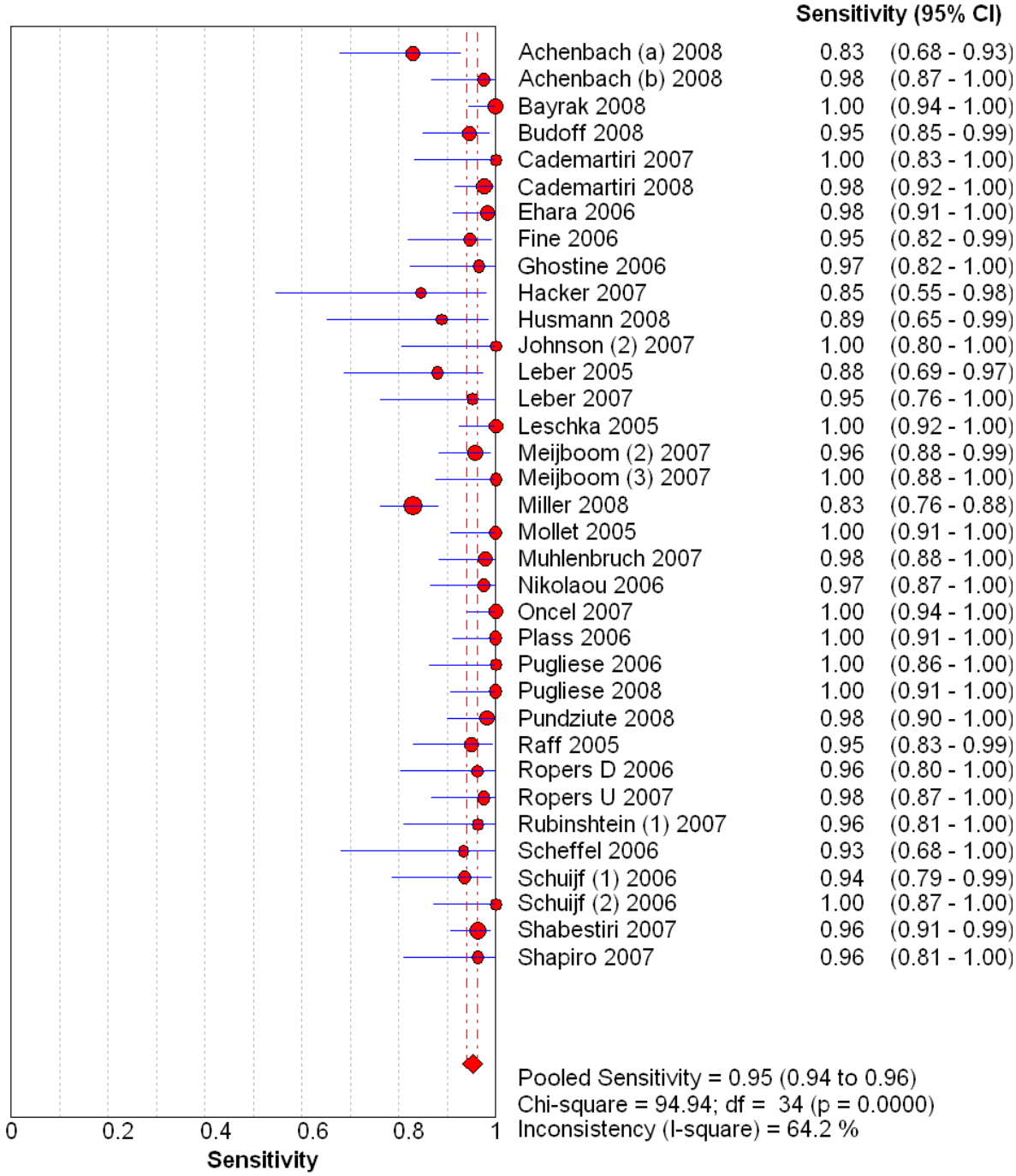
Age Group ≥62 years



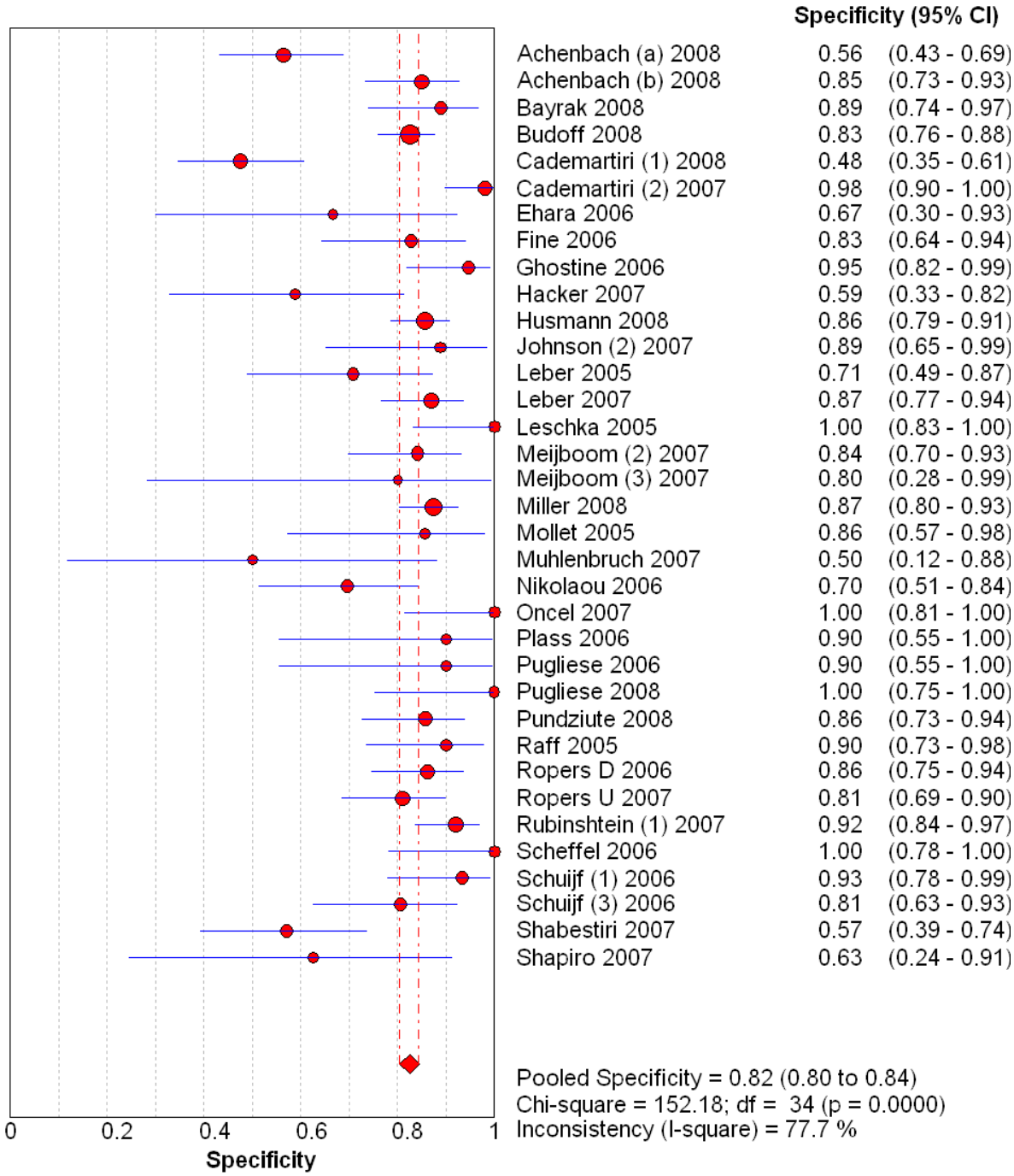
APPENDIX H:

RESULTS OF META-ANALYSES INCLUDING NEW MULTI-CENTER STUDIES

Sensitivity



Specificity



APPENDIX I:

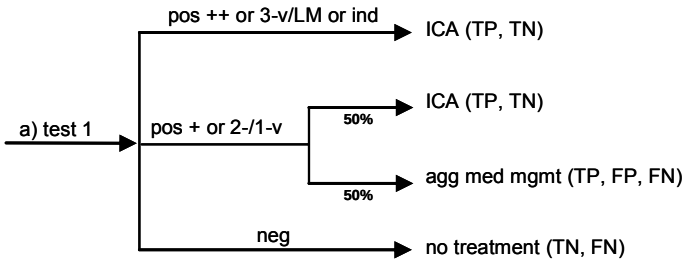
ADDITIONAL ANALYSES FROM ECONOMIC MODEL

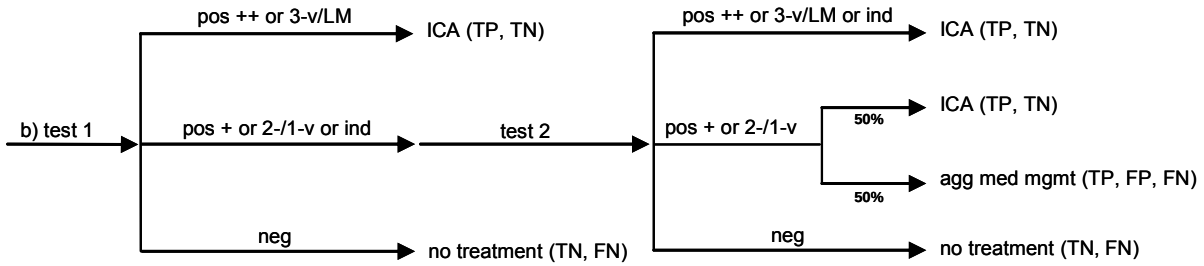
ED model: base case results using Stress ECHO for diagnostic workup in SOC

Outcomes (per 1,000)	SOC	CCTA + SOC
True positive	189	252
True negative	731	731
False negative	80	17
False negative results with ACS	22	5
Referred for ICA	389	323
ICA negative results	200	71
ICA related deaths	0.04	0.03
Incidental findings	0	138
Costs ($ per patient)		
ED/patient	1,099	1,377
Delay/patient	443	33
Cath lab/patient	1,070	888
Total/patient	**2,612**	**2,298**
Cost difference (CCTA vs. SOC)	**- $314**	

Notes: SOC: standard of care; CCTA: CCTA: coronary computed tomographic angiography

Outpatient model: diagnostic pathways "more aggressive treatment for mild stenosis"





Notes: pos ++: markedly abnormal test result, pos +: abnormal test result, ind: indeterminate results: TP: true-positive; TN: true-negative; FP: false-positive; FN: false-negative; ICA: invasive coronary angiography; agg med mgmt: aggressive medical management (according to AHA/ACC guidelines)

Outpatient model: diagnostic results – more aggressive treatment for mild-moderate stenosis (30 % CAD prevalence)

Estimates	CCTA	SPECT	SECHO	CCTA -> SPECT	SPECT -> CCTA	CCTA -> SECHO	SECHO -> CCTA	SECHO -> SPECT
True positive	288	271	245	266	265	245	239	228
False positive	44	76	38	12	14	6	10	18
True negative	661	629	667	693	691	699	694	687
False negative	8	25	50	29	31	51	56	68
FN w/3vd or LMd	2	1	4	2	1	2	4	4
Referred for ICA	249	318	300	200	191	193	171	187
ICA-negative results	63	133	125	18	17	16	12	28
ICA related deaths	0.25	0.32	0.30	0.20	0.19	0.19	0.17	0.19
Exposed to radiation	1000	1000	300	1000	1000	1000	408	408
Incidental findings requiring f/u	138	0	0	138	56	138	47	47
Total costs/patient [excluding all FU costs, $]	1,152	1,638	1,126	1,260	1,479	1,092	928	1,076

Outpatient model: diagnostic results – more aggressive treatment for mild-moderate stenosis (10 % CAD prevalence)

Estimates	CCTA	SPECT	SECHO	CCTA -> SPECT	SPECT -> CCTA	CCTA -> SECHO	SECHO -> CCTA	SECHO -> SPECT
True positive	96	91	82	89	89	81	80	75
False positive	56	97	47	14	18	7	13	22
True negative	845	804	854	887	883	894	888	879
False negative	3	8	17	10	10	18	19	24
FN w/3vd or LMd	1	0	1	1	0	1	1	1
Referred for ICA	146	235	223	85	83	81	72	91
ICA-negative results	83	171	163	25	23	23	17	38
ICA related deaths	0.15	0.23	0.22	0.09	0.08	0.08	0.07	0.09
Exposed to radiation	1000	1000	223	1000	1000	1000	292	292
Incidental findings requiring f/u	138	0	0	138	46	138	37	37
Total costs/patient [excluding all FU costs, $]	866	1,410	912	861	1,151	753	625	757

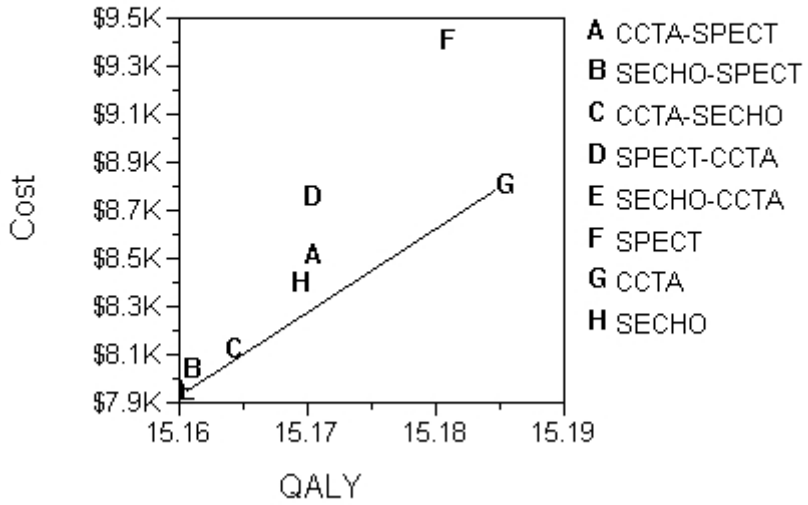
Outpatient model: Strategies ordered by increasing effectiveness (30% CAD prevalence) (more aggressive strategy of mild stenosis)

Strategy	Effectiveness (QALY)	Costs ($)
SECHO-CCTA	15.161	7,943
SECHO-SPECT	15.161	8,036
CCTA-SECHO	15.164	8,123
SECHO	15.169	8,400
CCTA-SPECT	15.170	8,517
SPECT-CCTA	15.170	8,754
SPECT	15.181	9,409
CCTA	15.185	8,802

Outpatient model: Incremental cost-effectiveness analysis (30% CAD prevalence) (more aggressive strategy of mild stenosis)

Strategy	Effect	Incr. Effect	Costs	Incr. Costs	Incr. C/E
SECHO	15.169		8,400		
CCTA	15.185	0.025	8,802	402	16,100

Outpatient model: Cost-effectiveness graph (30% CAD prevalence) (more aggressive strategy of mild stenosis)



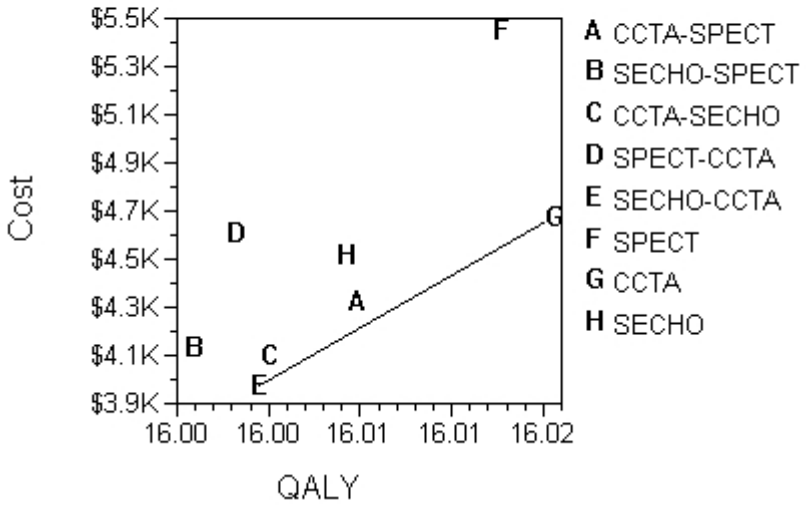
Outpatient model: Strategies ordered by increasing effectiveness (10% CAD prevalence) (more aggressive strategy of mild stenosis)

Strategy	Effectiveness (QALY)	Costs ($)
SECHO-SPECT	15.999	4,130
SPECT-CCTA	16.001	4,604
SECHO-CCTA	16.003	3,971
CCTA-SECHO	16.003	4,098
SECHO	16.007	4,516
CCTA-SPECT	16.008	4,321
SPECT	16.016	5,443
CCTA	16.019	4,668

Outpatient model: Incremental cost-effectiveness analysis (10% CAD prevalence) (more aggressive strategy of stenosis)

Strategy	Effect	Incr. Effect	Costs	Incr. Costs	Incr. C/E
SECHO	16.007		4,516		
CCTA	16.019	0.012	4,668	152	12,700

Cost-effectiveness graph (10% CAD prevalence) (more aggressive strategy of mild stenosis)



Diagnostic results based on CORE-64 diagnostic accuracy estimates (83% sens, 91% spec) (30 % CAD prevalence)

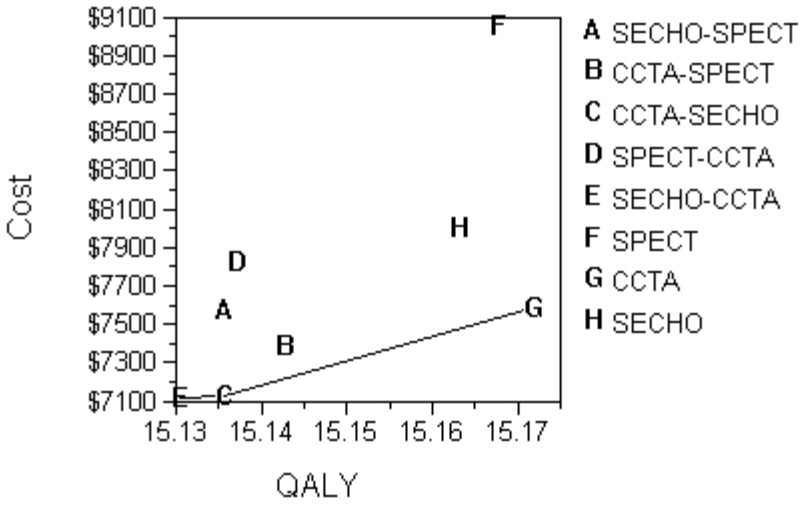
Estimates	CCTA	SPECT	SECHO	CCTA -> SPECT	SPECT -> CCTA	CCTA -> SECHO	SECHO -> CCTA	SECHO -> SPECT
True positive	245	271	245	226	237	207	215	228
False positive	60	149	74	16	18	8	13	33
True negative	645	556	631	689	687	697	691	672
False negative	50	25	50	70	58	88	81	68
False negative with 3vd or LM disease	13	1	4	13	2	14	6	4
Referred for ICA	95	160	195	89	89	99	84	105
ICA-negative results	21	61	89	5	5	8	4	12
ICA related deaths	0.10	0.16	0.20	0.09	0.09	0.10	0.08	0.11
Exposed to radiation	1000	1000	195	1000	1000	1000	342	1000
Incidental findings requiring f/u	138	0	0	138	56	138	47	36
Total costs/patient [excluding all FU costs, $]	728	1,204	837	911	1,201	815	690	850

Strategies ordered by increasing effectiveness (30% CAD prevalence) based on CORE-64 diagnostic accuracy estimates (83% sens, 91% spec)

Strategy	Effectiveness (QALY)	Costs ($)
SECHO-CCTA	15.135	7,120
SECHO-SPECT	15.140	7,576
CCTA-SECHO	15.140	7,130
SPECT-CCTA	15.141	7,827
CCTA-SPECT	15.147	7,382
SECHO	15.167	7,998
SPECT	15.172	9,051
CCTA	15.176	7,581

Notes: QALY: quality-adjusted life year

Cost-effectiveness graph based on CORE-64 diagnostic accuracy estimates (30% CAD prevalence) (83% sens, 91% spec)



HEALTH EVIDENCE REVIEW COMMISSION (HERC)

DRAFT COVERAGE GUIDANCE: MANAGEMENT OF RECURRENT ACUTE OTITIS MEDIA IN CHILDREN

DATE: XX/XX/XXXX

HERC COVERAGE GUIDANCE

Prophylactic antibiotics should be covered for recurrent otitis media.

Tympanostomy tubes may be covered for acute otitis media only for recurrent acute otitis media, defined as three or more episodes in six months or four or more episodes in one year.

Adenoidectomy or adenotonsillectomy should not be covered for the treatment of recurrent acute otitis media.

RATIONALE FOR GUIDANCE DEVELOPMENT

The HERC selects topics for guideline development or technology assessment based on the following principles:

- Represents a significant burden of disease
- Represents important uncertainty with regard to efficacy or harms
- Represents important variation or controversy in clinical care
- Represents high costs, significant economic impact
- Topic is of high public interest

Coverage guidance development follows to translate the evidence review to a policy decision. Coverage guidance may be based on an evidence-based guideline developed by the Evidence-based Guideline Subcommittee or a health technology assessment developed by the Heath Technology Assessment Subcommittee. In addition, coverage guidance may utilize an existing evidence report produced by one of HERC's trusted sources, generally within the last three years.

EVIDENCE SOURCES

Leach, A.J., & Morris, P.S. (2006). Antibiotics for the prevention of acute and chronic suppurative otitis media in children. *Cochrane Database of Systematic Reviews*, *4*(CD004401), 1-70. [Assessed as up-to-date: 5 AUG 2010]

McDonald, S., Langton Hewer, C.D., & Nunez, D.A. (2008). Grommets (ventilation tubes) for recurrent acute otitis media in children. *Cochrane Database of Systematic Reviews*, *4*(CD 004741), 1-14. [Assessed as up-to-date: 10 JAN 2011]

DRAFT

Shekelle PG, Takata G, Newberry SJ, Coker T, Limbos M, Chan LS, et al. (2010). *Management of Acute Otitis Media: Update.* Evidence Report/Technology Assessment No. 198. (Prepared by the RAND Evidence-Based Practice Center under Contract No. 290 2007 10056 I). Rockville, MD: Agency for Healthcare Research and Quality. Retrieved September 26, 2012, from http://www.ncbi.nlm.nih.gov/books/NBK56132/

The summary of evidence in this document is derived directly from these evidence sources, and portions are extracted verbatim.

SUMMARY OF EVIDENCE

Clinical Background

Acute Otitis Media (AOM) is a viral and/or bacterial infection of the middle ear and represents the most common childhood infection for which antibiotics are prescribed in the United States. A diagnosis of AOM requires 1) a history of acute onset of signs and symptoms, 2) the presence of middle ear effusion, and 3) signs and symptoms of middle-ear inflammation. There is a high rate of spontaneous resolution for AOM, but if left untreated it can occasionally lead to complications such as acute mastoiditis. The optimal duration of antibiotic therapy is not known and varies worldwide from none to 10 days. One recent strategy is to delay antibiotic treatment until symptoms persist or worsen after several days. Recurrent AOM is generally defined as three episodes in the previous six months or four episodes in the prior year, and has been treated with prophylactic antibiotics or pressure equalization tubes (PE tubes).

Evidence Review

Prevention of AOM in patients with recurrent OM – Medical therapy

The AHRQ review was unable to reach definitive conclusions regarding the comparative effectiveness of different antibiotics for AOM in children with recurrent otitis media. For recurrent otitis media, authors relied on an earlier version of the Leach systematic review to conclude that long-term antibiotic administration was found to decrease AOM episodes from 3 to 1.5 for every 12 months of treatment per otitis prone child during active treatment. The authors caution that the potential consequences of long-term treatment need to be considered.

A Cochrane review (Leach 2011) included 17 studies of children at increased risk of AOM. In seven of these, increased risk was defined as three episodes of AOM in the previous six months or four episodes in the previous year. The other studies defined high risk in a variety of ways, but most included prior episodes of AOM. All excluded children with immunodeficiency or craniofacial abnormalities. In this meta-analysis, long-term antibiotics reduced any episode of AOM and the number of episodes of AOM,

with approximately five children needing to be treated long-term to prevent one child experiencing AOM. Antibiotics prevented 1.5 episodes of AOM for every 12 months of treatment per child. Long-term antibiotics were not associated with a significant increase in adverse events.

Prevention of AOM in patients with recurrent OM – Surgical therapy

The Cochrane systematic review addressed the effectiveness of tympanostomy tubes in children with recurrent acute otitis media (defined as three or more acute infections in six months, or four or more acute infections in a year). It included only two randomized controlled trials (RCTs) with a total of 176 children (McDonald 2008). Both trials included children under age three who had a history of at least three episodes of AOM in the six months prior to referral. In one trial, the control was no treatment and in the other, it was daily sulfamethoxazole/trimethoprim syrup at 12mg/kg/day. Both trials reported results categorically as “no episodes of AOM” or “one or more episodes of AOM”, and both found that PE tubes reduce the occurrence of AOM at a follow up of six months, with the larger trial that used a no-treatment control reaching statistical significance. There was no follow up in either trial longer than six months, nor were any harms reported.

The AHRQ report included five RCTs that addressed adenoidectomy, with or without tonsillectomy or tympanostomy. One trial compared adenoidectomy to sulfafurazole and found no significant difference, although the trend was toward favoring the drug. Two trials compared adenoidectomy to placebo, and while both favored the procedure, neither reached statistical significance. The same was true for the trial that compared adenotonsillectomy to adenoidectomy alone; the trend favored adenotonsillectomy, but results did not reach statistical significance. When adenotonsillectomy was compared to placebo, there was 15% improvement in success rate (defined as no AOM episodes for one year), but given the wide confidence interval, this did not meet the required minimum clinically important difference of 5% adopted by the authors. Lastly, one trial compared adenoidectomy plus PE tubes to PE tubes alone, and found no difference between groups in number of episodes of AOM in the following year. Differences in harms, when reported, were either inconclusive or equivalent.

Overall Summary

For recurrent AOM, prophylactic antibiotics modestly decrease the number of episodes of AOM, with a number needed to treat of five. Pressure equalization tubes may reduce the frequency of acute otitis media in the short-term. Adenoidectomy does not result in a clinically significant decrease in the frequency of AOM.

DRAFT

PROCEDURE

Placement of pressure equalization tubes
Antibiotic Pharmacotherapy
Adenoidectomy
Adenotonsillectomy

DIAGNOSES

Acute otitis media
Recurrent acute otitis media

APPLICABLE CODES

CODES	DESCRIPTION
ICD-9 Diagnosis Codes	
381	Nonsuppurative otitis media and eustachian tube disorder
382	Suppurative and unspecified otitis media
315.34	Speech and language developmental delay due to hearing loss
389	Hearing loss
ICD-9 Volume 3 (Procedure Codes)	
None	
CPT Codes	
42820	Tonsillectomy and adenoidectomy; younger than age 12
42821	Tonsillectomy and adenoidectomy; age 12 and over
42830	Adenoidectomy, primary; younger than age 12
42831	Adenoidectomy, primary; age 12 and over
42835	Adenoidectomy, secondary; younger than age 12
42836	Adenoidectomy, secondary; age 12 and over
69433	Tympanostomy (requiring insertion of ventilating tube, local or topical anesthesia)
69436	Tympanostomy (requiring insertion of ventilating tube, general anesthesia)
69424	Ventilating tube removal requiring general anesthesia
HCPCS Codes	
None	

Note: Inclusion on this list does not guarantee coverage

DRAFT

Coverage guidance is prepared by the Health Evidence Review Commission (HERC), HERC staff, and subcommittee members. The evidence summary is prepared by the Center for Evidence-based Policy at Oregon Health & Science University (the Center). This document is intended to guide public and private purchasers in Oregon in making informed decisions about health care services.

The Center is not engaged in rendering any clinical, legal, business or other professional advice. The statements in this document do not represent official policy positions of the Center. Researchers involved in preparing this document have no affiliations or financial involvement that conflict with material presented in this document.

DRAFT

Preface

The Agency for Healthcare Research and Quality (AHRQ) conducts the Effective Health Care Program as part of its mission to organize knowledge and make it available to inform decisions about health care. As part of the Medicare Prescription Drug, Improvement, and Modernization Act of 2003, Congress directed AHRQ to conduct and support research on the comparative outcomes, clinical effectiveness, and appropriateness of pharmaceuticals, devices, and health care services to meet the needs of Medicare, Medicaid, and the Children's Health Insurance Program (CHIP).

AHRQ has an established network of Evidence-based Practice Centers (EPCs) that produce Evidence Reports/Technology Assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care. The EPCs now lend their expertise to the Effective Health Care Program by conducting comparative effectiveness reviews (CERs) of medications, devices, and other relevant interventions, including strategies for how these items and services can best be organized, managed, and delivered.

Systematic reviews are the building blocks underlying evidence-based practice; they focus attention on the strength and limits of evidence from research studies about the effectiveness and safety of a clinical intervention. In the context of developing recommendations for practice, systematic reviews are useful because they define the strengths and limits of the evidence, clarifying whether assertions about the value of the intervention are based on strong evidence from clinical studies. For more information about systematic reviews, see http://www.effectivehealthcare.ahrq.gov/reference/purpose.cfm

AHRQ expects that CERs will be helpful to health plans, providers, purchasers, government programs, and the health care system as a whole. In addition, AHRQ is committed to presenting information in different formats so that consumers who make decisions about their own and their family's health can benefit from the evidence.

Transparency and stakeholder input from are essential to the Effective Health Care Program. Please visit the Web site (http://www.effectivehealthcare.ahrq.gov) to see draft research questions and reports or to join an e-mail list to learn about new program products and opportunities for input. Comparative Effectiveness Reviews will be updated regularly.

We welcome comments on this CER. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by e-mail to epc@ahrq.hhs.gov.

Carolyn M. Clancy, M.D.
Director
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director and Task Order Officer
Evidence-based Practice Program
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Acknowledgments

The authors gratefully acknowledge the continuing support of our AHRQ Task Order Officers, Margaret Coopey, M.P.S., M.H.A., R.N., and Stephanie Chang, M.D., and that of the former Task Order Officer, Steven Fox, M.D. We extend our appreciation to our Key Informants and members of our Technical Expert Panel (<NAME>), all of whom provided thoughtful advice and input during our research process.

The investigators deeply appreciate the considerable support, commitment, and contributions of the EPC team staff at <NAME>. We express our gratitude to the following individuals for their contributions to this project: <NAME, degrees>

Key Informants

Key Informants are the end users of research, including patients and caregivers, practicing clinicians, relevant professional and consumer organizations, purchasers of health care, and others with experience in making health care decisions. Within the EPC program, the Key Informant role is to provide input into identifying the Key Questions for research that will inform health care decisions. The EPC solicits input from Key Informants when developing questions for systematic review or when identifying high priority research gaps and needed new research. Key Informants are not involved in analyzing the evidence or writing the report and have not reviewed the report, except as given the opportunity to do so through the peer or public review mechanism

Key Informants must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their role as end-users, individuals are invited to serve as Key Informants and those who present with potential conflicts may be retained. The Task Order Officer and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

Key Informants

To be completed after peer review.

Technical Expert Panel

Technical Experts comprise a multi-disciplinary group of clinical, content, and methodological experts who provide input in defining populations, interventions, comparisons, or outcomes as well as identifying particular studies or databases to search. They are selected to provide broad expertise and perspectives specific to the topic under development. Divergent and conflicted opinions are common and perceived as health scientific discourse that results in a thoughtful, relevant systematic review. Therefore study questions, design and/or methodological approaches do not necessarily represent the views of individual technical and content experts. Technical Experts provide information to the EPC to identify literature search strategies and recommend approaches to specific issues as requested by the EPC. Technical Experts do not do analysis of any kind nor contribute to the writing of the report and have not reviewed the report, except as given the opportunity to do so through the peer or public review mechanism

Technical Experts must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals are invited to serve as Technical Experts and those who present

with potential conflicts may be retained. The Task Order Officer and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

Technical Expert Panel

To be completed after peer review.

Peer Reviewers

Peer Reviewers are invited to provide written comments on the draft report based on their clinical, content, or methodological expertise. Peer review comments on the preliminary draft of the report are considered by the EPC in preparation of the final draft of the report. Peer Reviewers do not participate in writing or editing of the final report or other products. The synthesis of the scientific literature presented in the final report does not necessarily represent the views of individual reviewers. The dispositions of the peer review comments are documented and will, for CERs and Technical briefs, be published three months after the publication of the Evidence report.

Potential Reviewers must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Invited Peer Reviewers may not have any financial conflict of interest greater than $10,000. Peer Reviewers who disclose potential business or professional conflicts of interest may submit comments on draft reports through the public comment mechanism.

<Name>
<Place>
<City>, <ST>

Otitis Media with Effusion: Comparative Effectiveness of Treatments

Structured Abstract

Objectives. To compare benefits and harms currently in use for treating otitis media with effusion (OME). Treatment for OME may include single approaches alone or combinations of two or more approaches. We focused on the following interventions and comparisons among them: surgical (tympanostomy tubes, myringotomy, and adenoidectomy); nonpharmacological interventions (autoinflation); pharmacological (oral or nasal steroids); complementary and alternative medicine (CAM); and other treatment strategies (watchful waiting).

Data Sources. We identified five recent systematic reviews during topic refinement. We searched MEDLINE,® EMBASE,® the Cochrane Library, the Cumulative Index to Nursing and Allied Health Literature (CINAHL®), from root through January 8, 2012 for additional studies. We identified additional studies from reference lists and experts.

Review Methods. Two people independently selected, extracted data from, and rated the risk of bias of relevant studies. We incorporated meta-analyses from the earlier reviews. Because of clinical heterogeneity, insufficient numbers of similar studies, or insufficient or variation in outcome reporting, we synthesized the additional data qualitatively. Two reviewers graded the strength of evidence (SOE) using established criteria.

Results. We identified five systematic reviews and 20 additional studies meeting our inclusion criteria. Generally, studies examined interventions in non-infant pediatric populations. We found no evidence on CAM. Clinical outcomes (KQ 1) included OME signs and symptoms, measured hearing, episodes of acute otitis media (AOM) and vestibular function. We found high SOE that tympanostomy tubes decrease the presence of middle ear effusion in comparison to watchful waiting or myringotomy, for up to 2 years, and improve hearing for up to 9 months (moderate SOE). Hearing levels were no longer better with tympanostomy tubes between 7 to 12 months post surgery (measured by ears) (low SOE) and at 12 and 18 months post surgery (measured by child) (moderate SOE). OME was more likely to resolve in children after adenoidectomy than in those with no treatment at 6 and 12 month followup (high SOE). We found no differences in hearing at various time points for up to three years between tympanostomy tubes (with adenoidectomy) and myringotomy (with adenoidectomy) or between topical steroids and controls at three and nine months followup (low SOE). We found no evidence concerning vestibular function or health care utilization and only one small study examining AOM outcomes (insufficient SOE). Functional outcomes included speech and language comprehension, behavior and quality of life. We found low SOE for no difference in language comprehension and expression at 6 to 9 months followup between tympanostomy tubes and myringotomy. In relation to harms (KQ3), tympanosclerosis and otorrhea were more common in ears with tympanostomy tubes (low SOE). All evidence was insufficient comparing types of tympanostomy tubes and approaches to myringotomy. We were unable to disentangle results related to watchful waiting from myringotomy. Evidence on subgroups (KQ4) was limited to one study of children with sleep apnea and one of adults. We found no evidence on the comparative effective of interventions by any health care factors (KQ5).

Conclusions. Overall, we found uneven bodies of evidence across treatment comparisons and outcomes. Compared with watchful waiting or myringotomy alone, tympanostomy tubes decreased effusion through at least 2 years of followup and improved hearing through at least 6 months but tubes do not differ from these other treatments thereafter. However, tube placement also increases the rate of multiple side effects such as otorrhea and tympanosclerosis. There was not sufficient evidence to resolve this trade-off using functional outcomes, quality of life or long term outcomes. Nor was there evidence for determining which kind of tubes or tube routines might provide the best risk benefit ratio. However, there is evidence that adenoidectomy and myringotomy improve effusion and hearing more rapidly than myringotomy alone through 2 years post surgery. Tubes do not add any benefit to adenoidectomy in comparison to adenoidectomy plus myringotomy, and placement of tubes increases side effects. Additional research is needed to develop a sufficient evidence base to support treatment decisions, particularly in subpopulations.

Contents

Executive Summary ES-1

Introduction 1
Background 1
Anatomy and Cause of Otitis Media with Effusion 1
Populations With Otitis Media With Effusion 1
Symptoms of Otitis Media with Effusion 2
Diagnosis of Otitis Media with Effusion 2
Treatments for Otitis Media With Effusion 3
Scope and Key Questions 7
Scope 7
Organization of This Report 9
Methods 11
Topic Refinement and Protocol Review 11
Literature Search Strategy 11
Search Strategy 11
Inclusion and Exclusion Criteria 12
Population 12
Intervention 13
Comparators 13
Outcomes 13
Timing 13
Setting 13
Study Designs 13
Study Selection 15
Data Extraction 15
Risk of Bias Assessment 15
Data Synthesis 16
Strength of the Body of Evidence 16
Applicability 17
Peer Review and Public Commentary 18
Results 19
Introduction 19
Results of Literature Searches 20
KQ 1: Comparative Effectiveness of Interventions: Clinical Outcomes or Health Care Utilization 22
Surgical Interventions: Tympanostomy Tube Comparisons 22
Surgical Interventions: Tympanostomy Tubes Versus Watchful Waiting or Myringotomy 27
Surgical Interventions: Tympanostomy Tubes Plus Adenoidectomy Versus Myringotomy Plus Adenoidectomy or Adenoidectomy Alone 32
Surgical Interventions: Myringotomy Comparisons 37
Surgical Interventions: Myringotomy with Adenoidectomy Comparisons 39

Surgical Interventions: Adenoidectomy Versus Other Interventions 40
Pharmaceutical Interventions: Oral or Topical Nasal Steroids 44
Nonpharmaceutical Interventions: Autoinflation 49
KQ 2: Comparative Effectiveness of Interventions: Functional and Quality of Life Outcomes 52
Surgical Interventions: Tympanostomy Tube Comparisons 52
Surgical Interventions: Tympanostomy Tubes Versus Watchful Waiting or Myringotomy 52
Surgical Interventions: Tympanostomy Tubes Plus Adenoidectomy Versus Myringotomy Plus Adenoidectomy or Adenoidectomy Alone 55
Surgical Interventions: Myringotomy Comparisons 55
Surgical Interventions: Myringotomy Plus Adenoidectomy Comparisons 56
Surgical Interventions: Adenoidectomy Versus Other Interventions 56
Pharmaceutical Interventions: Oral or Topical Nasal Steroids 56
Nonpharmaceutical Interventions: Autoinflation 57
KQ 3: Harms or Tolerability 57
Surgical Interventions: Tympanostomy Tube Comparisons 57
Surgical Interventions: Tympanostomy Tubes Versus Watchful Waiting or Myringotomy 61
Surgical Interventions: Tympanostomy Tubes Plus Adenoidectomy Versus Myringotomy Plus Adenoidectomy or Adenoidectomy Alone 65
Surgical Interventions: Myringotomy Comparisons 66
Surgical Interventions: Myringotomy Plus Adenoidectomy Comparisons 67
Surgical Interventions: Adenoidectomy Versus Other Interventions 67
Pharmaceutical Interventions: Oral or Topical Nasal Steroids 67
Nonpharmaceutical: Autoinflation 68
KQ 4: Comparative Effectiveness of Interventions for Subgroups of Patients 68
Surgical Interventions: Tympanostomy Tube Comparisons 68
Surgical Interventions: Tympanostomy Tubes Versus Watchful Waiting or Myringotomy 68
Surgical Interventions: Tympanostomy Tubes Plus Adenoidectomy Versus Myringotomy Plus Adenoidectomy or Adenoidectomy Alone 68
Surgical Interventions: Myringotomy Comparisons 69
Surgical Interventions: Myringotomy With Adenoidectomy Comparisons 69
Surgical Interventions: Adenoidectomy Versus Other Interventions 69
Pharmaceutical Interventions: Oral or Topical Nasal Steroids 69
Nonpharmaceutical Interventions: Autoinflation 69
KQ 5: Comparative Effectiveness by Health Care Factors 70
Discussion 71
Key Findings and Strength of Evidence 71
Overview 71
KQ 1: Clinical Outcomes 72
KQ 2: Health-Related Quality Of Life and Functional Outcomes 74
KQ 3: Harms Associated With Interventions to Treat Otitis Media With Effusion 75
KQ 4: Outcomes for Important Patient Subgroups 76
KQ 5: Health Care Factors 77

Findings in Relation to What Is Already Known 77
Implications for Clinical and Policy Decisionmaking 77
Applicability 78
Population 78
Intervention and Comparators 79
Outcomes 79
Time Frames 79
Settings 80
Limitations of the Review Process 80
Limitations of the Evidence Base 80
Research Gaps 80
Gaps in Subgroups Studied 81
Gaps in Outcomes Measured (Benefits or Harms) 82
Gaps in Interventions 82
Deficiencies in Methods 84
Conclusions 84
References 85

Tables

Table A. Strength of evidence for interventions to improve clinical outcomes ES-13
Table B. Health-related quality of life and functional status ES-14
Table C. Strength of evidence for harms of interventions ES-15
Table D. Strength of evidence for subgroups ES-15

Table 1. Treatments for otitis media with effusion, with presumed mechanism of action 3
Table 2. Core PICOTS for this review 7
Table 3. Inclusion and exclusion criteria for studies of Otitis Media with Effusion 12
Table 4. Study inclusion criteria for review of otitis media with effusion 14
Table 5. Definitions of the grades of overall strength of evidence 17
Table 6. Description of procedures/measures of hearing, language and quality of life related to OME 20
Table 7. Characteristics of studies: Tympanostomy tube comparisons 22
Table 8. Clinical outcomes: Tympanostomy tube comparisons 25
Table 9. Characteristics of studies: Tympanostomy tubes versus watchful waiting or myringotomy 27
Table 10. Strength of evidence for KQ 1: Clinical outcomes and health care utilization 29
Table 11. Clinical outcomes: Tympanostomy tubes versus watchful waiting or myringotomy 30
Table 12. Characteristics of studies: Tympanostomy tubes and adenoidectomy versus watchful waiting or myringotomy and adenoidectomy or adenoidectomy alone 32
Table 13. Strength of evidence for KQ 1: Clinical outcomes and health care utilization 34
Table 14. Clinical outcomes: Tympanostomy tubes plus adenoidectomy versus watchful waiting or myringotomy plus adenoidectomy or adenoidectomy alone 35
Table 15. Characteristics of studies: Myringotomy comparisons 38
Table 16. Clinical outcomes: Myringotomy comparisons 38
Table 17. Characteristics of studies: Myringotomy with adenoidectomy 39
Table 18. Clinical outcomes: Myringotomy with adenoidectomy comparisons 40

Table 19. Characteristics of studies: Adenoidectomy versus other interventions 40
Table 20. Strength of evidence for KQ 1: Clinical outcomes and health care utilization 41
Table 21. Clinical outcomes: Adenoidectomy versus other interventions 42
Table 22. Characteristics of studies: Oral or topical nasal steroids 45
Table 23. Strength of evidence: Clinical outcomes and health care utilization 47
Table 24. Clinical outcomes: Oral or topical nasal steroids 48
Table 25. Characteristics of studies: Autoinflation 49
Table 26. Strength of evidence for KQ 1: Autoinflation 50
Table 27. Clinical outcomes: Autoinflation 51
Table 28. Strength of evidence for KQ 2: Tympanostomy tubes versus watchful waiting or myringotomy 52
Table 29. Functional and health-related quality-of-life outcomes: Tympanostomy tubes versus other treatments 53
Table 30. Functional and health-related quality-of-life outcomes: Tympanostomy tubes plus adenoidectomy versus myringotomy plus adenoidectomy or adenoidectomy alone 55
Table 31. Functional and health-related quality-of-life outcomes: Oral or topical nasal steroids 57
Table 32. Treatment harms or tolerability: Tympanostomy tube comparison studies 58
Table 33. Harms or tolerability: Tympanostomy tubes versus watchful waiting or myringotomy 62
Table 34. Harms or tolerability: Tympanostomy tubes plus adenoidectomy versus myringotomy plus adenoidectomy or adenoidectomy alone 66
Table 35. Treatment harms or tolerability: Oral or topical nasal steroids 67
Table 36. Treatment harms or tolerability: Autoinflation 68
Table 37. Comparative effectiveness for adults: Autoinflation 70
Table 38. Strength of evidence for interventions to improve clinical outcomes 73
Table 39. Health-related quality of life and functional status 75
Table 40. Strength of evidence for harms or tolerability of interventions 76
Table 41. Strength of evidence for subgroups 77

Figures

Figure A. Analytic framework for comparisons ES-3
Figure B. Disposition of articles ES-7

Figure 1. Analytic framework for review of treatments of otitis media with effusion 9
Figure 2. Disposition of articles 21

Appendixes

Appendix A. Search Strategy
Appendix B. Excluded Studies
Appendix C. Evidence Tables
Appendix D. Abstract and Full Text Forms
Appendix E. Risk of Bias Tables
Appendix F. Strength of Evidence Tables
Appendix G. Glossary
Appendix H. List of Acronyms

Executive Summary

Background

Definition of OME

Otitis media with effusion (OME) is defined as a collection of fluid in the middle ear without signs or symptoms of ear infection.[1] It has been thought that OME develops when negative pressure develops in the middle ear relative to atmospheric pressure and then fluid accumulates because of that pressure.[2] However, various other potential explanations include ciliary dysfunction, proliferation of fluid producing goblet cells, allergy and residual bacterial antigens, and biofilm.[3] The presence of fluid in the middle ear then decreases tympanic membrane and middle ear function, leading to decreased hearing, a 'fullness' sensation in the ear, and occasionally pain from the pressure changes.

Prevalence of OME

OME occurs commonly during childhood, with as many as 90 percent of children (80% of individual ears) having at least one episode of OME by age 10.[4] Some subpopulations of children are disproportionately affected by OME. Those with cleft palate, Down syndrome, and other craniofacial anomalies are at high risk for anatomic causes of OME and compromised function of the Eustachian tube.[5] Individuals of American Indian, Alaskan, and Asian backgrounds are believed to be at greater risk,[6] as are children with adenoid hyperplasia. In addition, children with sensorineural hearing loss will likely be more affected by the secondary conductive hearing loss that occurs with OME.

Although rare, OME also occurs in adults. This usually happens after patients develop a severe upper respiratory infection such as sinusitis, severe allergies, or rapid change in air pressure after a plane flight or a scuba dive. The incidence of prolonged OME in adults is not known, but it is much less common than in children.[7]

Many episodes of OME resolve spontaneously within 3 months, but 30 to 40 percent of children have recurrent episodes and 5 to 10 percent of cases last more than 1 year.[1, 8, 9]

Diagnosis of OME

Diagnostically, the core feature of OME is middle ear effusion (MEE)—i.e., sticky or thick fluid behind the eardrum in the middle ear. Tympanocentesis, which is the removal of fluid from behind the eardrum by using a needle to puncture the tympanic membrane, remains the gold standard for diagnosing MEE and OME. A variety of supplemental examination techniques assist with identification. The most studied additional diagnostic method is pneumatic otoscopy, which is considered an accurate way to diagnose MEE by trained examiners.[4] To use this procedure, clinicians blow air through the otoscope, causing movement of the tympanic membrane that they can then compare with normal movement of the membrane. Tympanometry is a supplemental diagnostic tool that indirectly measures middle ear pressure and tympanic membrane mobility. A "flat" tympanogram (Type B tympanogram) is consistent with OME. Additionally, children with OME often have a corresponding conductive hearing loss on pure-tone audiometry.

Types of Treatments

Given the natural history of OME including spontaneous resolution in most patients over time, clinical decisions are complicated. Despite recent practice guidelines and systematic reviews,[5, 10-17] the comparative benefits and harms of treatments and treatment strategies for OME are uncertain. Surgical procedures, medications, and alternative treatments have all been used to treat OME.

During topic refinement, The Evidence-based Practice Center (EPC) considered each of the known treatments in terms of uncertainty within the published literature (including gaps in the evidence), clinical importance, patient important outcomes, and relevance to the U.S. population. The set of treatments examined in this review are indicated under Key Question 1.

Scope and Key Questions

The EPC was charged with conducting this review because of the continuing uncertainty about efficacy, effectiveness, and particularly comparative effectiveness, as well as harms, for the included therapies. Providing more up-to-date and comprehensive comparative information will be helpful to many stakeholder groups in making decisions about when and how to treat this condition. This comparative review includes all interventions currently in use for treating OME—surgical, pharmacological, and nonpharmacological—except antibiotics, antihistamines and decongestants which have been extensively reviewed previously and demonstrated to have no benefit in this population.

The intent of the review was to cover the entire range of individuals who have OME, including populations who have not been examined in past reviews such as adults and children with special conditions such as Down syndrome, cleft palate, or existing hearing loss. We did not limit the timeframe for outcomes nor did we exclude any settings.

The EPC addressed five key questions (KQs) in this comparative effectiveness review.

KQ1: What is the comparative effectiveness of the following treatment options (active treatments and watchful waiting) in affecting clinical outcomes or health care utilization in patients with OME? Treatment options include: tympanostomy tubes, myringotomy, oral or topical nasal steroids, autoinflation, complementary and alternative medical (CAM) procedures, watchful waiting, and variations in surgical technique or procedures

KQ2: What is the comparative effectiveness of the different treatment options listed in KQ1 (active treatments, watchful waiting, and variations in surgical procedures) in improving functional and health-related quality of life outcomes in patients with OME?

KQ3: What are the harms or tolerability among the different treatment options?

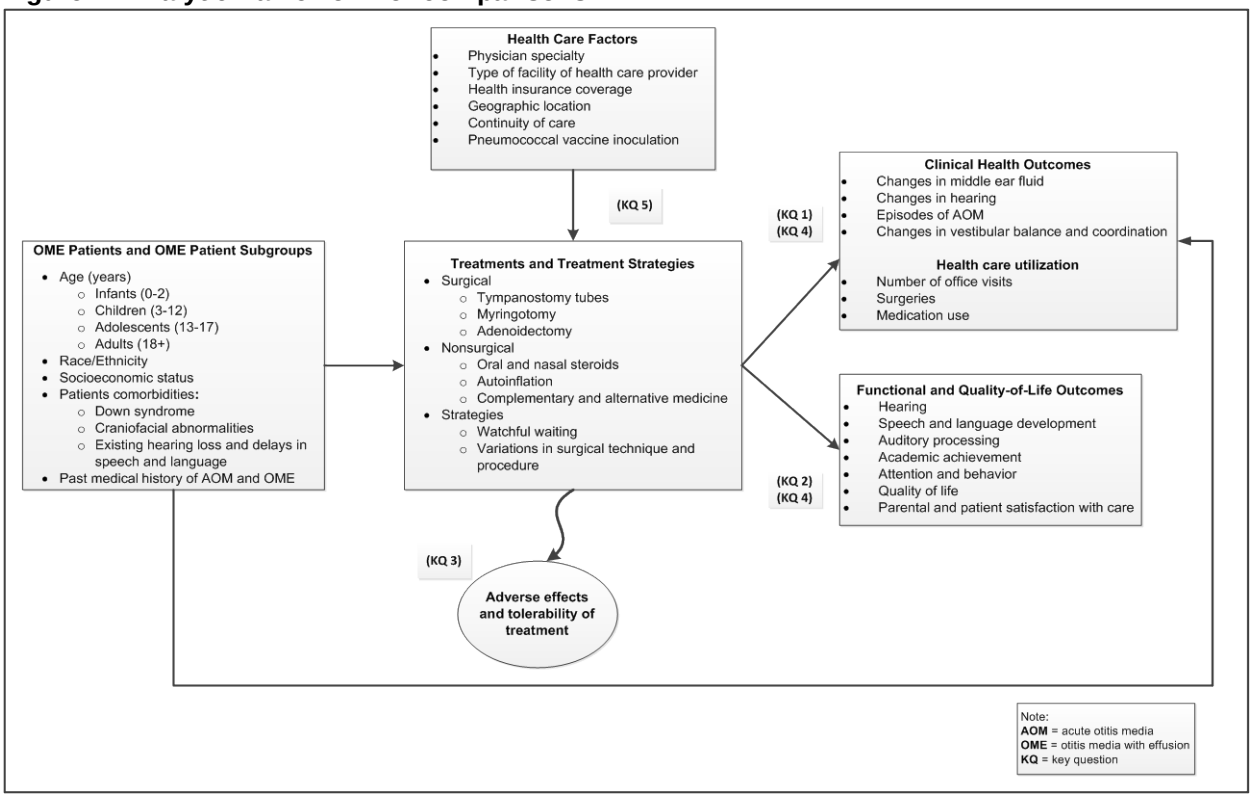
KQ4: What are the comparative benefits and harms of treatment options in subgroups of patients with OME?

KQ5: Is the comparative effectiveness of treatment options related to factors affecting health care delivery or the receipt of pneumococcal vaccine inoculation?

We developed an analytic framework (Figure A) for this review. The populations of interest are in the box to the far left; the interventions appear in the middle; and the two sets of outcomes

(for KQ1 and KQ2 on benefits, and also KQ4 on important subgroups) appear on the far right. KQ3 concerns harms (various types of adverse events). Finally, KQ5 relates to a set of health care delivery or clinical factors (pneumococcal vaccination) that may influence choices of treatments or their clinical and quality-of-life outcomes.

Figure A. Analytic framework for comparisons



Methods

Literature Search Strategy

Search Strategy

Five recently published systematic reviews on comparisons of interest were identified during the topic refinement stage of the review. As discussed in our review protocol, The Cochrane Collaboration conducted four of them, and the Swedish Council of Technology in Health Care commissioned the fifth. The reviews covered the following treatment options for OME: tympanostomy tubes, adenoidectomy, steroids, and autoinflation. To avoid duplicating the work of these teams, we used these reviews as a starting point. We limited our review and analysis for each of our KQs to evidence that these systematic reviews provided plus added evidence that these recent reviews did not consider (i.e., observational studies done at any time, newer studies published since the last search dates in those reviews, and studies focusing on populations excluded from the reviews), such as adults with OME or children with Down syndrome or cleft palate, who may be differently affected by OME.

We searched MEDLINE® (via PubMed), EMBASE, ® the Cochrane Library, and the Cumulative Index to Nursing and Allied Health Literature (CINAHL®) to identify studies not

included in the systematic reviews. An experienced research librarian used a predefined list of search terms and medical subject headings (MeSH). We reviewed our search strategy with our Technical Expert Panel (TEP) and incorporated their input into our search strategy. We limited the electronic searches to English language materials, based on limited resources. We completed the initial search on 1/8/2012, and we will complete an update during peer review.

We searched unpublished and grey literature relevant to the review topic. Methods for identifying grey literature included a review of trial registries. In addition, AHRQ requested Scientific Information Packets (SIPs) from the developers and distributors of the interventions identified in the literature review. We included unpublished studies that met all inclusion criteria and contained enough information on their research methods to permit us to make a standard risk-of-bias assessment of individual studies. Finally, we manually searched reference lists of reviews, including trials and background articles, to look for relevant citations that our searches might have missed and that addressed our KQs. We imported all citations into an electronic database (EndNote® X4).

Inclusion and Exclusion Criteria

We developed inclusion and exclusion criteria with respect to the PICOTS framework. The review only included English language studies of individuals with OME. We included the five systematic reviews identified during topic refinement and eligible studies not included in those reviews included randomized controlled trials (RCTs), nonrandomized controlled trials (NRCTS) and cohort studies. There were no other restrictions, which permitted studies that included individuals of any age, racial/ethnic background, or co-morbidity.

The treatments of interest were tympanostomy tubes, myringotomy, adenoidectomy, oral or intranasal steroids, autoinflation of the Eustachian tube, CAM procedures, watchful waiting, and variations in surgical technique or procedures. With two exceptions, included studies had to compare at least two of these treatments. We considered inactive controls in comparison with steroid treatment and usual care in comparison with autoinflation, based on the Cochrane Collaboration systematic review inclusion criteria.

We specified a broad range of outcomes (see Figure A). We included clinical outcomes such as changes in middle ear fluid, episodes of AOM and hearing thresholds; health care utilization; functional and quality of life outcomes such as speech and language development, behavior, and parental satisfaction with care; and harms.

We were primarily interested in treatment outcomes of 3 months or longer but included outcomes of less than 3 months. We focused on end-of-intervention results when they were the only endpoint data available, such as in the autoinflation treatment studies. .

We did not exclude studies based on geography or the setting of the service provision.

Study Selection

Two people independently reviewed article abstracts using the inclusion/exclusion criteria. If both reviewers agreed that the study did not meet eligibility criteria, we excluded it; otherwise, two reviewers again independently reviewed the full-text article. If both reviewers agreed that a study did not meet eligibility criteria, we excluded it. Each reviewer recorded the primary reason for exclusion. If the reviewers disagreed about the primary reason for exclusion, they resolved conflicts by discussion and consensus or by consulting a third member of the team. We screened unpublished studies identified through a grey literature search and review of SIPs using the same title/abstract and full-text review processes.

Data Abstraction

We developed a template for evidence tables for data synthesis using the PICOTS framework. For the five systematic reviews and additional studies that met our inclusion criteria, we abstracted relevant information into these evidence tables. We abstracted the following information as presented in the systematic reviews and from the additional studies: characteristics of study populations, interventions, comparators, settings, study designs, methods, and results. We only reviewed individual studies included in the systematic reviews to determine the availability of subgroup analyses not included in the reviews. We did not reabstract the articles included in the systematic reviews. One trained reviewer initially abstracted the relevant data from each included article and a second member of the team reviewed each data abstraction against the original article for completeness and accuracy. All data abstraction was performed using Microsoft Excel® software.

Risk of Bias Assessment

For each included study, we assessed the potential for selection bias, performance bias, attrition bias, detection bias, and outcome reporting bias using an instrument designed to evaluate risk of bias in a trial or observational study and a second designed to evaluate the risk of bias of a systematic review that have been successfully used previously at our EPC. We did not reevaluate the risk of bias of the studies included in the systematic reviews but all had been determined to be low or medium risk of bias by the original review study authors. Two independent reviewers rated the risk of bias for each study. Disagreements between the two reviewers were resolved by discussion and consensus or by consulting a third member of the team. Results of this assessment are summarized in a rating of low, medium, or high risk of bias. High risk of bias studies were those that had at least one major issue that had the potential to cause significant bias and might invalidate the results.

Data Synthesis

Across all included studies, the populations, interventions, and outcome measures in the additional data were heterogeneous and did not lend themselves to a pooled analysis. They also did not lend themselves to updating the meta-analyses from the five earlier systematic reviews. Because we determined that quantitative analyses were not appropriate, we did all analyses qualitatively. Evidence used in the synthesis included the results from the earlier meta-analyses, additional data from individual studies contained in those systematic reviews, and data from the articles included from our own searches.

Strength of the Body of Evidence

We graded the strength of evidence (SOE) based on the guidance established for the AHRQ Effective Health Care Program EPCs conducting comparative effectiveness reviews, as detailed in the paper by Owens and colleagues.[18] The EPC approach incorporates four key domains: risk of bias, consistency, directness, and precision of the evidence. The overall grade for strength of evidence is based on the scores for the four domains and reflects the strength of the body of evidence to answer the KQs on the comparative effectiveness, efficacy, and harms of the treatments and treatment strategies covered in this review.

A grade of high SOE indicates that we have high confidence that the evidence reflects the true effect. Moderate SOE implies that we have moderate confidence that the evidence reflects

the true effect. Low SOE suggests that we have low confidence that the evidence reflects the true effect. Insufficient SOE signifies either that evidence is completely unavailable or that it does not permit estimation of an effect.

Two reviewers assessed each domain independently and also assigned an overall grade for each key outcome listed in the framework; they resolved any conflicts through consensus discussion. If they did not reach consensus, the team brought in a third party to settle the conflict.

Applicability

We assessed the applicability of individual studies as well as the body of evidence. For individual studies, we examined factors that may limit applicability based on the PICOTS structure such as population characteristics, intervention characteristics, and comparators. We abstracted key characteristics of applicability into the evidence tables. During data synthesis, we assessed the applicability of the body of evidence using the abstracted characteristics. KQ4 includes a detailed analysis of intervention efficacy in population subgroups.

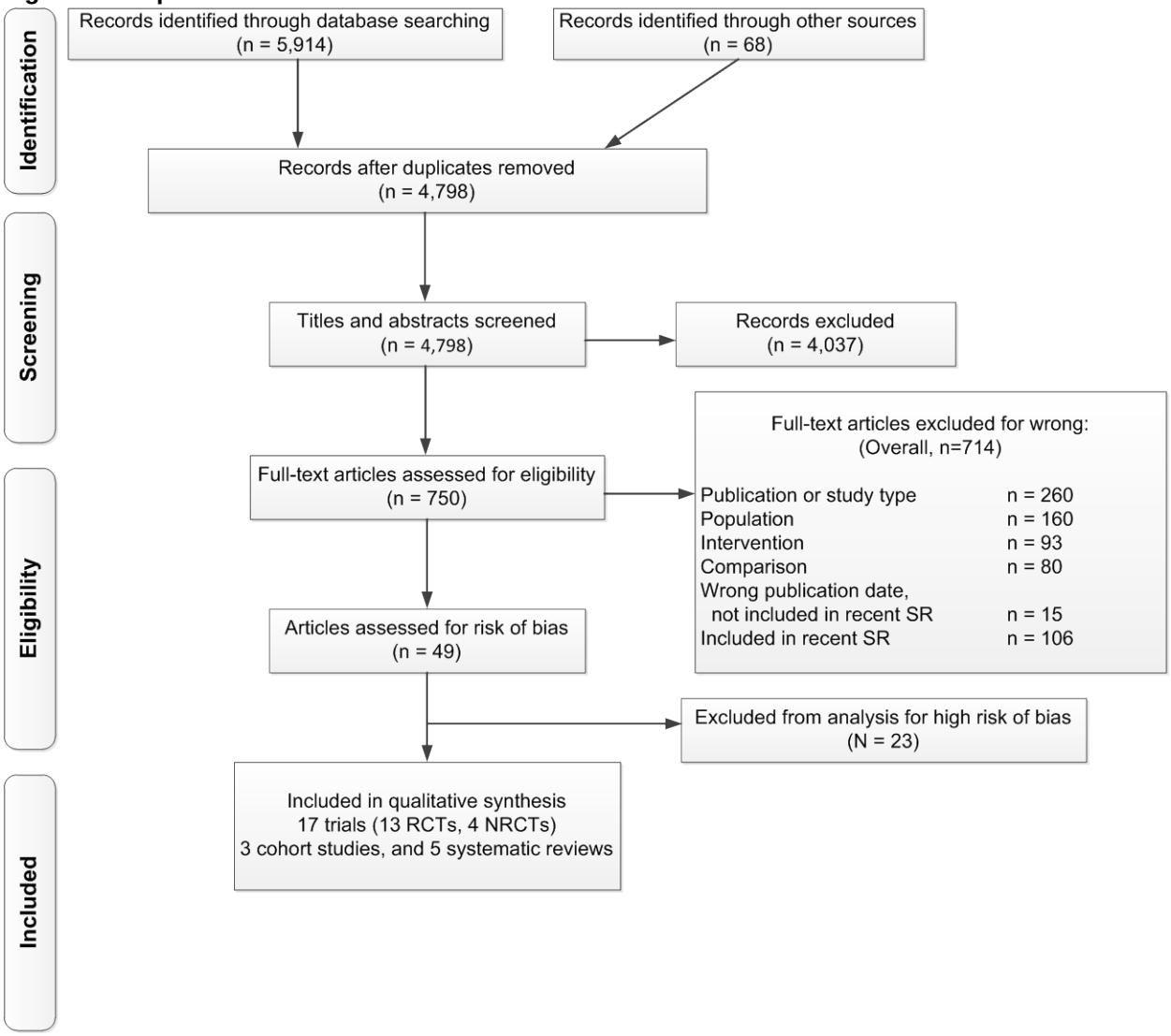
Results

This section is organized by KQ and then grouped by intervention comparison. The summaries of evidence findings are presented in Tables A-D by KQ. Summary tables can be found in the full report. Evidence tables of included studies may be found in Appendix C, and the strength of evidence ratings for the main outcomes of each KQ are detailed in Appendix F.

Literature Searches

We identified a total of 4798 unduplicated citations and determined that 750 met criteria for full-text review (Figure B). We excluded 701 full-text articles based on our inclusion criteria and before risk of bias assessment. We further eliminated 23 articles because of high risk of bias. Overall, we included 21 full-text articles, detailing 20 studies (13 RCTs, four nonrandomized controlled trials, and three observational studies) and five systematic reviews. We recorded the reason that each excluded full-text publication failed to satisfy the eligibility criteria and compiled a comprehensive list of such studies (Appendix B of the full report). We did not include 23 high risk of bias studies in our analyses (Appendix C of the full report). Of these, six were RCTs, seven were non-randomized trials and the remainder were cohort studies. Virtually all lacked information on any baseline patient characteristics; of particular concern, unknown differences between groups based on age or time with OME could invalidate outcomes. Other serious concerns were a lack of control for selective concurrent treatment and lack of control for confounders in cohort studies.

Figure B. Disposition of articles



Abbreviations: NRCT = non-randomized controlled trial; RCT = randomized controlled trial; SR = systematic review

Key Question 1. Comparative Effectiveness: Clinical Outcomes or Health Care Utilization

Eight of the 20 studies and four of the five systematic reviews reported differences in one or more indices of OME signs and symptoms as a function of treatment. Two of the individual studies and three of the systematic reviews reported differences in hearing as a function of treatment. No studies reported health utilization or balance outcomes. Evidence for decrease in subsequent AOM was insufficient as only one study examined this outcome.

Tympanostomy Tube Comparisons

Seven individual studies[20-26] and seven studies from one systematic[10] review addressed tympanostomy tube comparisons. The seven individual studies included four RCTs,[20-22, 25] one nonrandomized controlled trial[26] and two observational studies.[23, 24] The Hellstrom et al. review[10]

contained seven studies, five of which were RCTs and two of which were nonrandomized controlled trials. These single studies compared different types of tubes (e.g., materials, size), approaches to insertion, or topical prophylaxis therapies. All comparisons were made between ears of the same individual. Primarily because of the diversity of comparisons, the evidence is insufficient for middle ear status or hearing.

Tympanostomy Tubes Versus Watchful Waiting/Myringotomy

Two of the individual studies that met our inclusion criteria[27, 28] and two systematic reviews[10, 12] addressed comparisons between trials of tympanostomy tubes with either myringotomy or watchful waiting. Browning et al.[12] reviewed 10 studies, eight of which were in comparison to watchful waiting or delayed treatment and two were in comparison to myringotomy in the control ear. Hellstrom et al.[10] included six of the studies that were in the Browning review. Both of the individual studies we reviewed and all of the studies included in the systematic reviews were RCTs.

Tympanostomy tube placement decreased time with middle ear effusion through 2 years post surgery (high [1 year] to moderate [2 year] SOE); evidence was insufficient for longer followup. Thereafter, OME decreased in watchful waiting and myringotomy groups as well. In contrast, tympanostomy tubes only improved hearing through 9 months post surgery (high [4 to 6 months] to moderate [6 to 9 months] SOE). Thereafter the differences in hearing became attenuated and were not significant at 7 to 12 months (low SOE) and 12 to 18 months post surgery (moderate SOE).

Tympanostomy Tubes Plus Adenoidectomy Versus Myringotomy Plus Adenoidectomy or Adenoidectomy Alone

We identified eight studies examining outcomes in relation to tympanostomy tubes plus adenoidectomy as compared to myringotomy plus adenoidectomy or adenoidectomy alone.[29-36] Three of the studies compared tympanostomy tubes in one ear to a control ear in children who all had adenoidectomies. One of the studies was an RCT,[29] and the other two were nonrandomized controlled studies.[30, 31] The other five studies compared tympanostomy tubes to myringotomy, among children who all had adenoidectomies; four were RCTs,[32-35] and one was a nonrandomized control study.[36]

The evidence was insufficient for examining OME signs and symptoms, with only two single studies examining middle ear effusion and recurrence of OME. We found no differences in hearing at any endpoint in five studies between tympanostomy tubes and myringotomy among children who also received adenoidectomies (low SOE). We found mixed results for tubes compared with watchful waiting in children who also received adenoidectomies (insufficient SOE).

Myringotomy Comparisons

Only one RCT compared two different procedures for myringotomy on both middle ear and hearing outcomes.[37] The two procedures were radio frequency myringotomy with Mitomycin C, a topical chemotherapeutic agent and radio frequency myringotomy alone. A majority of individuals in each arm received adenoidectomy (73% and 67% respectively). There was insufficient evidence for concluding superiority of either myringotomy procedure for OME signs and symptoms or hearing outcomes.

Myringotomy Plus Adenoidectomy Comparisons

One retrospective cohort study compared two different procedures for myringotomy.[38] The comparison was between laser myringotomy and cold knife myringotomy. In both groups all individuals received an adenoidectomy. The evidence is insufficient for determining superiority for either myringotomy approach for OME signs and symptoms. No study examined hearing or any other clinical outcome.

Adenoidectomy Versus Other Interventions

One Cochrane Collaboration systematic review by van den Aadweg et al.[13] provided all the evidence for adenoidectomy in comparison to tympanostomy tubes, myringotomy, or no surgery. The review included seven RCTs that were limited to OME patients. The trials examined adenoidectomy with and without myringotomy versus non-surgical treatment or myringotomy only; adenoidectomy with unilateral tympanostomy tubes versus a unilateral tympanostomy tube only (comparison by ears); and adenoidectomy with bilateral tympanostomy tubes versus bilateral tympanostomy tubes only.

Adenoidectomy was superior to no treatment for resolution of OME at both 6 months (risk difference of 0.27 measured through otoscopy and 0.22 as measured through tympanometry; high SOE) and 12 months post surgery (risk difference of 0.29 through tympanometry; high SOE). One single study found that adenoidectomy and myringotomy were superior to myringotomy alone for reducing time with effusion and improving hearing at 24 months (low SOE). There was insufficient evidence to determine the effectiveness of adenoidectomy as compared with no treatment, myringotomy, or tympanostomy tubes for AOM.

Oral or Topical Nasal Steroids

We included one RCT[39, 40] that examined topical intranasal steroids and one recent systematic review conducted by Cochrane Collaboration[15] that examined oral steroids and topical intranasal steroids. Simpson et al.[15] summarized the evidence from nine RCTs of oral steroids and two RCTs of topical intranasal steroids. Both the systematic review and the study we reviewed examined signs and symptoms of OME and hearing. Meta-analyses[15] comparing oral steroids with controls with and without antibiotics did not show differences in middle ear effusion at either 1-2 months post treatment (moderate SOE). There was insufficient evidence comparing oral steroids with controls for hearing at any time point. The RCT[39, 40] comparing intranasal steroids with controls did not find differences in OME cure rate or in hearing at one or more months post treatment (low SOE). There was insufficient evidence for comparing either oral or topical intranasal steroids with controls for any other clinical outcome.

Autoinflation

One Cochrane Collaboration systematic review conducted by Pera et al.[14] summarized evidence from six RCTs of any form of autoinflation, a technique designed to increase pressure in the oropharynx forcing open the Eustachian tube though a nasal balloon or other process. The review included five studies with children and one study with adults, 16-75 years of age. All studies were in comparison to no autoinflation, and other treatments (e.g., antibiotics, analgesics) were permitted as long as they were given equally to both arms. Meta-analyses comparing autoinflation with controls found an improvement in OME at 1 month or less, post treatment

(low SOE). Evidence was insufficient for drawing conclusions regarding improvements in OME at longer time periods or for other clinical outcomes, including hearing.

Key Question 2. Comparative Effectiveness: Functional Outcomes or Quality of Life

Only a subset of the treatment comparisons reported functional or quality of life outcomes. These include tympanostomy tubes versus watchful waiting, tympanostomy tubes plus adenoidectomy versus myringotomy plus adenoidectomy, and steroids versus control. In general, there were no differences between the treatments. The studies that are included are described under KQ1.

Tympanostomy Tubes Versus Watchful Waiting/Myringotomy

Meta-analysis reported by Browning et al.[12] did not find any differences in language development at 6 and 9 months post treatment between tubes and watchful waiting (low SOE). Individual studies included in the systematic review reported mixed findings for behavior outcomes (insufficient SOE). Evidence was insufficient for determining differences between tubes and watchful waiting for quality of life outcomes. No studies comparing tubes with myringotomy reported on functional or quality of life outcomes (insufficient SOE).

Tympanostomy Tubes Plus Adenoidectomy Versus Myringotomy Plus Adenoidectomy

One study comparing tubes plus adenoidectomy with myringotomy plus adenoidectomy reported quality of life outcomes.[35] The two groups did not differ at any time point (insufficient SOE). Strength of evidence was insufficient for all speech/language, cognitive, and behavioral outcomes because there were no studies including these outcomes.

Oral or Topical Nasal Steroids

One small study included in the systematic review[15] and the one clinical trial that met our inclusion criteria [39, 40] examined functional outcomes. Patients receiving intranasal steroids plus oral antibiotics did not differ in parents' assessment of their children's symptoms from patients receiving placebo plus antibiotics (low SOE); nor did patients receiving intranasal steroids differ from controls in parent reported hearing outcomes (low SOE). No studies comparing topical or oral steroids to control examined any other functional outcomes (insufficient SOE).

Key Question 3. Harms or Tolerability

Five of the treatment comparisons included in the review reported on harms. These include tube versus tube studies, tubes versus watchful waiting/myringotomy, tubes plus adenoidectomy versus myringotomy plus adenoidectomy/adenoidectomy alone, steroids, and autoinflation. Only a limited range of harms were included for any comparison. Few significant differences in harms were reported.

Tympanostomy Tube Comparisons

We reviewed five studies that reported on otorrhea[20] and one systematic review Hellstrom et al.[10] reported on three other studies that examined otorrhea. We found insufficient evidence for one type of tube or tube routine because the results were mixed. The other harms that were

reported in one or more studies include repeat tube placement, cholesteatoma, occlusion, tympanosclerosis, and the presence of granulation tissue. The results were reported in only one study, were non significant or were conflicting (insufficient SOE).

Tympanostomy Tubes Versus Watchful Waiting/Myringotomy

The two studies we reviewed[27, 28] and both systematic reviews[10, 12] that compared tympanostomy tubes to watchful waiting/myringotomy examined harms. These included otorrhea, acute AOM, atrophy, tympanosclerosis, perforation, cholesteatoma, and the presence of granulation tissue. In comparison to either ears with no surgery or myringotomy, ears with tympanostomy tubes were more likely to experience tympanosclerosis (low SOE) or otorrhea (low SOE). Evidence was insufficient for other harms due to either conflicting results or having been reported in only a single study.

Tubes Plus Adenoidectomy Versus Myringotomy Plus Adenoidectomy/Adenoidectomy Alone

We reviewed seven studies that examined harms.[29, 31-36] These included repeat tubes, otorrhea, perforation, and tympanosclerosis or myringosclerosis. Results were either mixed or were reported in single studies (insufficient SOE)

Oral or Topical Nasal Steroids

Both the study we reviewed[39, 40] and the systematic review[15] reported on harms. The trials[39, 40] found no significant differences in mild adverse harms such as stinging nose, nose bleed, dry throat or cough between those receiving nasal steroids and those receiving placebo control (low SOE). Evidence concerning serious harms was sparse for either nasal or oral steroids (insufficient SOE).

Autoinflation

The one systematic review that compared autoinflation to control[14] provided no quantitative information on rates of serious or mild harms, only verbal statements indicating that there were few harms noted (insufficient SOE).

Key Question 4: Comparative Effectiveness of Interventions for Subgroups of Patients

Two treatment comparisons examined comparative effectiveness of interventions for subgroups of patients –tympanostomy tubes plus adenoidectomy versus myringotomy plus adenoidectomy/adenoidectomy alone and autoinflation.

Tubes plus Adenoidectomy Versus Myringotomy plus Adenoidectomy or Adenoidectomy Alone

One study that met our inclusion criteria[35] included children with sleep apnea and OME. The study did not report differences in hearing thresholds between children who received tubes plus adenoidectomy and children who received myringotomy plus adenoidectomy. Quality of life scores were inconsistent (insufficient SOE).

Autoinflation

One study included in the systematic review of autoinflation[14] included adults 16 to 75 years of age. The autoinflation group was significantly more likely to experience a complete recovery than those in the control group at the end of treatment and 50 days later (low SOE).

Key Question 5: Comparative Effectiveness by Health Care Factors

No included studies or systematic reviews examined effectiveness of intervention comparisons by any health care factors.

Discussion

Key Findings and Strength of Evidence

Key Question 1. Comparative Effectiveness: Clinical Outcomes or Health Care Utilization

Table A summarizes the SOE for comparative effectiveness of treatments on clinical outcomes. We are able to draw some conclusions regarding surgical treatments. We found high strength of evidence that in comparison to myringotomy or watchful waiting, tympanostomy tubes decreased the number of children with MEE at 1 year post surgery. Tubes continued to improve middle ear effusion at 2 years but the effect was observed less frequently. Tubes also improved hearing relative to watchful waiting but the effect was shorter in duration. We found only limited evidence for drawing conclusions about the relative benefits of tubes for other clinical outcomes such as OME recurrence or episodes of OME. We examined the evidence for whether tubes or myringotomy differentially improved clinical outcomes when they were added to adenoidectomy. Based on finding no differences in hearing at any time point in five studies, we concluded that there is low strength of evidence of no difference in hearing and insufficient evidence for any other clinical outcomes. We found high strength of evidence for concluding that adenoidectomy is superior to either tubes or no treatment for improving the likelihood of OME resolution at 6 and 12 months post surgery. We found low strength of evidence for adenoidectomy and myringotomy's superiority in relation to myringotomy alone at 2 years post surgery for improving OME resolution and hearing. Evidence was insufficient for other outcomes. All evidence concerning comparisons of tympanostomy tubes in relation to design or routes or techniques of insertion was evaluated as insufficient because there were only single studies that tested specific comparisons. Evidence was also insufficient for comparisons between different approaches to myringotomy with and without adenoidectomy due to the limited number of studies.

Table A. Strength of evidence for interventions to improve clinical outcomes

Intervention and Comparator	Number of Studies (Sample sizes)	Outcome and Results	Strength of Evidence
Tympanostomy tubes vs. watchful waiting or myringotomy	MA of 3 RCTs (N=574)	Tympanostomy tubes decreased persistent middle ear effusion at 1 year: 32% less time (95% CI, 17% to 48%)	High for benefit
	MA of 3 RCTs (N=426)	Tympanostomy tubes decreased persistent middle ear effusion at 2 years: 13% less time (95% CI, 8% to 17%)	Moderate for benefit
	MA of 3 RCTs (by ears) (N=230)	Tympanostomy tube groups had better measured hearing at 4-6 months: -10dB (95% CI, -19.12 to -1.05)	High for benefit
	MA of 3 RCTs (N=523)	Tympanostomy tube groups had better measured hearing at 6-9 months: -4.20dB (95% CI, -4.00 to -2.39)	Moderate for benefit
	MA of 3 RCTs (by ears) (N=234)	No difference between groups in measured hearing at 7-12 months: -5.18dB (95% CI, -10.43 to 0.07)	Low for no difference
	MA of 2 RCTs (N=328)	No difference between groups in measured hearing at 12 months: -0.41dB (95% CI, -2.37 to 1.54)	Moderate for no difference
	MA of 2 RCTs (N=283)	No difference in measured hearing between groups at 18 months: -0.02 dB (95% CI, -3.22 to 3.18)	Moderate for no difference
Tympanostomy tubes + adenoidectomy vs. myringotomy + adenoidectomy	5 studies: 2 RCTs by person (N=130), 2 RCTs (by ears) (N=85), 1 NRCT (by ears) (N=146)	No difference in measured hearing between groups at 6 and 12 months and greater than 3 years.	Low for no difference
Adenoidectomy vs. tympanostomy tubes or myringotomy/no surgery	MA of 2 RCTs (N=153); MA of 3 RCTs (N=297)	Adenoidectomy better OME resolution at 6 months. The risk difference was 0.27 (95% CI, 0.13 to 0.42), measured through otoscopy and 0.22 (95% CI, 0.12 to 0.32) measured through tympanometry.	High for benefit
	MA of 3 RCTs (N=298)	Adenoidectomy better OME resolution at 12 months. The risk difference was 0.29 (95% CI, 0.19 to 0.39).	High for benefit
	1 study (N=237)	Adenoidectomy and myringotomy less mean time with effusion than myringotomy alone at 24 months: -0.76 standard mean difference (95% CI, -1.02 to -0.49)	Low for benefit
	1 study (N=237)	Adenoidectomy and myringotomy better hearing than with myringotomy alone at 24 months: -0.65 standard mean difference time with hearing level ≥ 20 worse ear (95% CI, -0.91 to -0.39)	Low for benefit
Oral steroids vs. controls	MA of 3 RCTs (N=106)	No difference in persisting OME at 1-2 months, (no antibiotics provided in either group) OR=0.55 (95% CI, 0.21 to 1.48)	Low for no difference
Oral steroids vs. control	MA of 3 RCTs (N=243)	No difference in persisting OME at 1-2 months (antibiotics provided to both groups) OR=0.75 (95% CI, 0.45 to 1.27)	Moderate for no difference
Topical intranasal steroids vs. controls	1 RCT (N=217)	No difference in OME cure rates at 1, 3, and 9 months	Low for no difference
	1 RCT (N=217)	No difference in hearing loss at 3 and 9 months	Low for no difference
Autoinflation vs. control	MA of 2 RCTs (N=185)	Improvement in OME at ≤ 1month RR=3.84 (tympanometry change C2 to C1 or A) and RR=2.72 (tympanometry change B to C1 or A)	Low for benefit

Abbreviations: CI = confidence intervals; dB = decibels; NRCT = non-randomized controlled trial; MA = meta-analysis; N = number; OME= otitis media with effusion; OR = odds ratio; RCT = randomized controlled trial; vs.= versus

We have reached some conclusions for nonsurgical interventions. We found low strength of evidence for concluding that oral steroids do not offer any improvements in OME at 1 to 2 months post treatment. Similarly, we found moderate evidence for concluding that oral steroids with antibiotics do not provide improvements in OME at 1 to 2 months. A recent low risk of bias study provided additional evidence that OME and hearing outcomes were not improved through

the use of topical intranasal steroids through 9 months post treatment. These findings support the current clinical practice guidelines against the use of oral and intranasal steroids in treating OME in children. Although we found low evidence that autoinflation improves middle ear effusion at less than 1 month post treatment, there was insufficient evidence for reaching conclusions for other outcomes because outcomes across studies testing autoinflation were not measured at consistent lengths of followup or through consistent measures.

KQ 2. Health-Related Quality Of Life and Functional Outcomes

Table B summarizes the SOE for health related quality of life and functional outcomes. We can provide limited evidence regarding these outcomes. We found that language comprehension and language expression outcomes at 6-9 months were not significantly better among children with OME who received tympanostomy tubes than among those who were limited to watchful waiting or myringotomy (strength of evidence low). Evidence was insufficient to reach conclusions related to differences in either behavioral or quality of life outcomes for this treatment comparison. Quality of life outcomes were measured in one small study comparing tympanostomy tubes and adenoidectomy versus myringotomy and adenoidectomy, but we considered the evidence to be insufficient to reach conclusions. We found that evidence was low for concluding that topical steroids do not improve parent reported hearing difficulties up to 9 months. However, evidence was insufficient to reach conclusions about other quality of outcomes for oral steroids.

Table B. Health-related quality of life and functional status

Intervention and Comparator	Number of Studies (Sample sizes)	Outcome and Results	Strength of Evidence
Tympanostomy tubes vs. watchful waiting or myringotomy	MA of 3 RCTs (N=394)	No difference in language comprehension at 6-9 months: Mean difference=0.09 (95% CI, -0.21 to 0.39)	Low for no difference
	MA of 3 RCTs (N=393)	No difference in language expression at 6-9 months: Mean difference=0.03 (95% CI, -0.41 to 0.49)	Low for no difference
Topical steroids vs. controls	1 study (N=144)	No difference in parent reported hearing difficulties at 3 and 9 months and median days with hearing loss at 3 months.	Low for no difference

Abbreviations: CI = confidence interval; MA = meta-analysis; N = number; vs. = versus; RCT = randomized control trial

KQ3. Harms Associated With Interventions to Treat Otitis Media with Effusion

Table C summarizes the OME interventions on which we had low, moderate or high strength of evidence for harms outcomes. Based on findings from five studies, we determined that we had low strength of evidence for concluding that tympanosclerosis was more common in children who had tympanostomy tubes than those who were actively monitored or who had myringotomy. In addition, we also had low evidence for concluding that otorrhea occurred more frequently in ears with tubes than those with myringotomy or no surgery. The results from the one study that met our inclusion criteria lead us to conclude that mild adverse events are not significantly higher through the use of topical nasal steroids than placebo with low strength of evidence for no difference. However, evidence was insufficient to reach conclusions related to oral steroids and serious adverse events from oral or topical steroids. We found no evidence concerning harms

from adenoidectomy, including no data on any risks from having a surgical procedure. Similarly, we found insufficient evidence concerning the surgical risks from the insertion of tympanostomy tubes or myringotomy procedures with adenoidectomy.

Table C. Strength of evidence for harms of interventions

Intervention and Comparator	Number of Studies (Sample sizes)	Outcome and Results	Strength of Evidence
Tympanostomy tubes vs. watchful waiting or myringotomy	5 studies (N=1129)	Tympanosclerosis occurred more frequently in ears that had tympanostomy tubes, based on examinations after the tubes had been extruded.	Low for harms
	4 studies (N=960)	Otorrhea occurred more frequently in ears with tympanostomy tubes.	Low for harms
Tympanostomy tubes plus adenoidectomy vs. adenoidectomy alone or with myringotomy	2 studies (N=237)	Tympanosclerosis occurred more frequently in ears with tympanostomy tubes.	Low for harms
Topical nasal steroids vs. control	1 study (N=170)	No difference in mild adverse events such as nasal stinging, dry throat, and cough	Low for no difference

Abbreviations: N = number; NR = not reported; SR = systematic review; vs. = versus;

KQ4. Outcomes for Important Patient Subgroups

Table D provides the limited evidence we found for patient subgroups. Although we had planned to examine treatment efficacy or harms for key subgroups characterized by clinical or sociodemographic factors (such as age), we could not identify studies that covered most of our subgroups of interest. One study examined children with sleep apnea and OME, and one examined adults with OME. Among children with sleep apnea, all of whom had adenoidectomy to treat that condition, tympanostomy tubes and myringotomy did not differ significantly in terms of any measured outcomes. The study of autoinflation that was included in the systematic review[14] found differences in rates of recovery between those receiving autoinflation and those who were in the control group. Individuals in the autoinflation group were significantly more likely to experience a complete recovery than those in the control group at both the end of treatment ($p<0.001$) and at 50 days after treatment ($p<0.001$). Similarly, the ears of the participants receiving autoinflation had better recovery rates than control ears at both time points ($p<0.001$). Evidence was low for benefit.

Table D. Strength of evidence for subgroups

Intervention and Comparator	Number of Studies (Sample sizes)	Subgroup and Results	Strength of Evidence
Tympanostomy tubes + adenoidectomy vs. myringotomy + adenoidectomy	1 RCT (N=52)	Children with sleep apnea and OME: no difference between tympanostomy tubes and myringotomy outcomes (all children had received adenoidectomies).	Insufficient (one study)
Autoinflation vs. control	1 RCT (N=198)	Adults (16-75) with OME: differences between groups in composite measure of recovery (otoscopy, tympanometry, audiometry) at end of tx and 50 days after tx.	Low for benefit (one study)

Abbreviations: OME = otitis media with effusion; tx = treatment; RCT= randomized control trial

KQ5: Health Care Factors

We found no studies that examined issues related to health insurance coverage, physician specialty, type of facility of the provider, geographic location of patients, presence or absence of

continuity of care, or prior use of pneumococcal virus inoculation. Evidence is thus insufficient for all such factors.

Applicability

This review was intended to apply to individuals with OME of all ages. Findings about all interventions are likely to be applicable to non-infant, otherwise healthy children. However, the evidence base is quite limited for adults and for infant children or children with major coexisting or congenital conditions, such as those with cleft palate, Down syndrome, and sensorineural hearing loss, who may be disproportionately affected by OME. We provided evidence on all the commonly used treatments for OME, including tympanostomy tubes, myringotomy, adenoidectomy and watchful waiting as well as steroids upon the advice of our TEP. We provided evidence for autoinflation, an alternative non-invasive treatment strategy. We also sought to include CAM procedures, but no studies met our inclusion criteria. We did not limit the outcomes of interest. However, the bulk of the literature concerned reductions in OME, hearing, and a variety of harms that were not uniformly included in all studies. We did not limit the time frame for followup. Although most studies measured outcomes between 3 and 12 months post treatment, for any given comparison time frames were rarely uniform. Studies were conducted in clinical settings. They generally included populations from the United States and Western Europe, but there were a few from other countries including Japan, Egypt, and Iran.

Research Gaps

Research gaps in treatments for OME exist in several areas. We recommend the following for improving the research base.

The first area is to expand research in subgroups that were targeted in this review but for whom there was no evidence. This includes infants and toddlers who are developmentally vulnerable for language acquisition and for whom a mild conductive loss over a shorter period of time can be more detrimental than for older children. Children with craniofacial anomalies such as cleft palate and other developmental disorders including Down syndrome and sensorineural hearing loss have not been a part of most treatment studies. Additionally, there is only limited research on treatment efficacy in adults as we were only able to identify one study that concerned treatments for adults.

The second area is to examine treatments that have heretofore not been subjected to rigorous procedures. For instance, despite the interest in CAM treatments, there is a lack of carefully designed investigations of these treatments. There are treatments being designed to counteract the ontological effects of gastroreflux disease; further research of promising treatments is welcome.

The third area in which we found a gap is methodological. There is a lack of uniformity in measures, lack of functional outcomes, time points for collecting outcomes, and inclusion of baseline measures. Few studies provide effect sizes and fail to report their statistical power. Missing data are often not addressed, and even if attrition is acknowledged, statistical procedures are rarely used to correct for this. Attention in this area would greatly improve the literature base.

Conclusions

Overall, we found uneven bodies of evidence across treatment comparisons and outcomes. Compared with watchful waiting or myringotomy alone, tympanostomy tubes decreased effusion

through at least 2 years of followup and improved hearing through at least 6 months. However, tube placement also increases the rate of multiple side effects such as otorrhea and tympanosclerosis. There was a low level of evidence that tube placement is not different than watchful waiting for functional outcomes, quality of life or long term outcomes. Adenoidectomy decreases the number of children with OME in the short term relative to watchful waiting. For children who are receiving adenoidectomy during an intervention, there is moderate evidence that the addition of tube placement does not improve outcomes compared with myringotomy alone. Additional research is needed to develop a comprehensive evidence base to support decision making among the various treatment options, particularly in subpopulations.

References

1. Stool SE, Berg AO, Berman S, et al. Otitis Media With Effusion in Young Children. Agency for Health Care Policy and Research Clinical Practice Guideline, No. 12. Rockville, MD: 1994. AHCPR Publication No. 94-0622.
2. Cantekin EI, Bluestone CD. Membrane ventilating tube for the middle ear. Ann Otol Rhinol Laryngol. 1976 Mar-Apr; 85(2 Suppl 25 Pt 2):270-6. PMID: 1267358.
3. Klein JO, Pelton S. Otitis media with effusion (serous otitis media) in children. In: Basow DS, ed UpToDate. Waltham, MA; 2012.
4. Shekelle P, Takata G, Chan LS, et al. Diagnosis, History, and Late Effects of Otitis Media With Effusion. Agency for Healthcare Research and Quality Evidence Report/Technology Assessment No. 55 (Prepared by Southern California Evidence-based Practice Center under Contract No 290-97-0001, Task Order No. 4). Rockville, MD: 2003. AHRQ Publication No. 03-E023.
5. National Collaborating Centre for Women's and Children's Health. Surgical Management of Otitis Media With Effusion in Children. NICE Clinical Guideline 60. London: National Institute for Health and Clinical Excellence; February 2008.
6. Maw AR. Otitis media with effusion (glue ear). In: Kerr AG, Groves J, Evans NG, eds. Scott-Browan's Otolaryngology. 5th ed. London: Butterworths; 1987:159-76.
7. Finkelstein Y, Ophir D, Talmi YP, et al. Adult-onset otitis media with effusion. Arch Otolaryngol Head Neck Surg. 1994 May; 120(5):517-27. PMID: 8172703.
8. Tos M. Epidemiology and natural history of secretory otitis. Am J Otol. 1984 Oct; 5(6):459-62. PMID: 6542752.
9. Williamson IG, Dunleavey J, Bain J, et al. The natural history of otitis media with effusion--a three-year study of the incidence and prevalence of abnormal tympanograms in four South West Hampshire infant and first schools. J Laryngol Otol. 1994 Nov; 108(11):930-4.
10. Hellstrom S, Groth A, Jorgensen F, et al. Ventilation tube treatment: a systematic review of the literature. Otolaryngol Head Neck Surg. 2011 Sep; 145(3):383-95. PMID: 21632976.
11. Williamson I. Otitis media with effusion in children. Clinical Evidence. 2011; pii: 0502PMID: 21477396.
12. Browning GG, Rovers MM, Williamson I, et al. Grommets (ventilation tubes) for hearing loss associated with otitis media with effusion in children. Cochrane Database Syst Rev. 2010; (10):CD001801. PMID: 20927726.
13. van den Aardweg MT, Schilder AG, Herkert E, et al. Adenoidectomy for otitis media in children. Cochrane Database Syst Rev. 2010; (1):CD007810. PMID: 20091650.
14. Perera R, Haynes J, Glasziou P, et al. Autoinflation for hearing loss associated with otitis media with effusion. Cochrane Database Syst Rev. 2006; (4):CD006285. PMID: 17054290.
15. Thomas CL, Simpson S, Butler CC, et al. Oral or topical nasal steroids for hearing loss associated with otitis media with effusion in children. Cochrane Database Syst Rev. 2006; (3):CD001935. PMID: 16855980.
16. Griffin GH, Flynn C, Bailey RE, et al. Antihistamines and/or decongestants for otitis media with effusion (OME) in children. Cochrane Database Syst Rev. 2006; (4):CD003423. PMID: 17054169.
17. Rosenfeld RM, Culpepper L, Doyle KJ, et al. Clinical practice guideline: Otitis media with effusion. Otolaryngol Head Neck Surg. 2004 May; 130(5 Suppl):S95-118. PMID: 15138413.
18. Owens DK, Lohr KN, Atkins D, et al. AHRQ series paper 5: grading the strength of a body of evidence when comparing medical interventions--agency for healthcare research and quality and the effective health-care program. J Clin Epidemiol. 2010 May; 63(5):513-23. PMID: 19595577.

19. Owens DK, Lohr KN, Atkins D, et al. AHRQ series paper 5: grading the strength of a body of evidence when comparing medical interventions--agency for healthcare research and quality and the effective health-care program. J Clin Epidemiol. 2010 May; 63(5):513-23. PMID: 19595577.

20. Ovesen T, Felding JU, Tommerup B, et al. Effect of N-acetylcysteine on the incidence of recurrence of otitis media with effusion and re-insertion of ventilation tubes. Acta Otolaryngol Suppl. 2000; 543:79-81. PMID: 10908985.

21. Wielinga EW, Smyth GD. Comparison of the Goode T-tube with the Armstrong tube in children with chronic otitis media with effusion. J Laryngol Otol. 1990 Aug; 104(8):608-10. PMID: 2230552.

22. McRae D, Gatland DJ, Youngs R, et al. Aspiration of middle ear effusions prior to grommet insertion an etiological factor in tympanosclerosis. J Otolaryngol. 1989 Aug; 18(5):229-31. PMID: 2769837.

23. Iwaki E, Saito T, Tsuda G, et al. Timing for removal of tympanic ventilation tube in children. Auris Nasus Larynx. 1998 Dec; 25(4):361-8. PMID: 9853658.

24. Slack RW, Gardner JM, Chatfield C. Otorrhoea in children with middle ear ventilation tubes: a comparison of different types of tubes. Clin Otolaryngol Allied Sci. 1987 Oct; 12(5):357-60. PMID: 3427799.

25. Licameli G, Johnston P, Luz J, et al. Phosphorylcholine-coated antibiotic tympanostomy tubes: are post-tube placement complications reduced? Int J Pediatr Otorhinolaryngol. 2008 Sep; 72(9):1323-8. PMID: 18635268.

26. Abdullah VA, Pringle MB, Shah NS. Use of the trimmed Shah permavent tube in the management of glue ear. J Laryngol Otol. 1994 Apr; 108(4):303-6. PMID: 8182314.

27. Koopman JP, Reuchlin AG, Kummer EE, et al. Laser myringotomy versus ventilation tubes in children with otitis media with effusion: a randomized trial. Laryngoscope. 2004 May; 114(5):844-9. PMID: 15126741.

28. Mandel EM, Rockette HE, Bluestone CD, et al. Myringotomy with and without tympanostomy tubes for chronic otitis media with effusion. Arch Otolaryngol Head Neck Surg. 1989 Oct; 115(10):1217-24. PMID: 2789777.

29. Brown MJ, Richards SH, Ambegaokar AG. Grommets and glue ear: a five-year follow up of a controlled trial. J R Soc Med. 1978 May; 71(5):353-6. PMID: 650647.

30. Austin DF. Adenotonsillectomy in the treatment of secretory otitis media. Ear Nose Throat J. 1994 Jun; 73(6):367-9, 73-4. PMID: 8076535.

31. Lildholdt T. Unilateral grommet insertion and adenoidectomy in bilateral secretory otitis media: preliminary report of the results in 91 children. Clin Otolaryngol Allied Sci. 1979 Apr; 4(2):87-93. PMID: 455774.

32. D'Eredità R, Shah UK. Contact diode laser myringotomy for medium-duration middle ear ventilation in children. Int J Pediatr Otorhinolaryngol; 2006. p. 1077-80.

33. Popova D, Varbanova S, Popov TM. Comparison between myringotomy and tympanostomy tubes in combination with adenoidectomy in 3-7-year-old children with otitis media with effusion. Int J Pediatr Otorhinolaryngol. 2010 Jul; 74(7):777-80. PMID: 20399511.

34. Shishegar M, Hoghoghi H. Comparison of adenoidectomy and myringotomy with and without tube placement in the short term hearing status of children with otitis media with effusion: a preliminary report. Iranian Journal of Medical Sciences; 2007. p. 169-72.

35. Vlastos IM, Houlakis M, Kandiloros D, et al. Adenoidectomy plus tympanostomy tube insertion versus adenoidectomy plus myringotomy in children with obstructive sleep apnoea syndrome. J Laryngol Otol. 2011 Mar; 125(3):274-8. PMID: 21205368.

36. Tos M, Stangerup SE. Hearing loss in tympanosclerosis caused by grommets. Arch Otolaryngol Head Neck Surg. 1989 Aug; 115(8):931-5. PMID: 2751852.

37. Ragab SM. The effect of radiofrequency and mitomycin C on the closure rate of human tympanostomy. Otol Neurotol. 2005 May; 26(3):355-60. PMID: 15891633.

38. Szeremeta W, Parameswaran MS, Isaacson G. Adenoidectomy with laser or incisional myringotomy for otitis media with effusion. Laryngoscope. 2000 Mar; 110(3 Pt 1):342-5. PMID: 10718416.

39. Williamson I, Benge S, Barton S, et al. Topical intranasal corticosteroids in 4-11 year old children with persistent bilateral otitis media with effusion in primary care: double blind randomised placebo controlled trial. BMJ. 2009; 339:b4984. PMID: 20015903.

40. Williamson I, Benge S, Barton S, et al. A double-blind randomised placebo-controlled trial of topical intranasal corticosteroids in 4- to 11-year-old children with persistent bilateral otitis media with effusion in primary care. Health Technol Assess. 2009 Aug; 13(37):1-144. PMID: 19671372.

Introduction

Background

Otitis media with effusion (OME) is defined as a collection of fluid in the middle ear without signs or symptoms of ear infection.[1] OME has been known by a variety of terms including serous otitis media, chronic otitis media, secretory otitis media, nonsuppurative otitis media, mucoid otitis media, or fluid in the middle ear.[1] This occurs commonly during childhood, with as many as 90 percent of children (80% of individual ears) who will have at least one episode of OME by age 10.[2] Medical treatment of these effusions costs an estimated $4 billion annually.[2] Despite the high prevalence, indications for treatment with definitive, low-risk treatments and an understanding of the impact of prolonged OME on individual health outcomes have been elusive.

Anatomy and Cause of Otitis Media with Effusion

The cascade of physiologic changes that lead to OME are initiated with dysfunction of the Eustachian tube. Normally this tube aerates the middle ear by connecting it to the posterior pharynx. The function of the Eustachian tube becomes evident during atmospheric ascent or descent. The pressure sensation that is experienced when taking off in an airplane comes from middle ear barometric pressure changes. The "popping" of the ear after that represents opening of the orifice to the Eustachian tube in the posterior pharynx and equalization of pressure between atmospheric pressure and the barometric pressure in the middle ear.

Clinicians do not fully understand the pathophysiology of OME. The traditional teaching has been that OME develops when a negative pressure develops in the middle ear relative to atmospheric pressure and then exudative or transudative fluid accumulates because of that pressure.[3] However, various other potential explanations include ciliary dysfunction, proliferation of fluid producing goblet cells, allergy and residual bacterial antigens, and biofilm.[4] The presence of fluid in the middle ear decreases tympanic membrane and middle ear function, leading to decreased hearing, a 'fullness' sensation in the ear, and occasionally pain from the pressure changes.

Several predisposing environmental factors are associated with an increased risk of developing OME.[5] These include being exposed to secondhand smoke, attending child care, and having environmentally induced allergies.

Populations With Otitis Media With Effusion

OME is typically considered an affliction of childhood. Many episodes of OME resolve spontaneously within 3 months, but 30 to 40 percent of children have recurrent episodes and 5 to 10 percent of cases last more than 1 year.[1, 6, 7]

Some subpopulations of children are disproportionately affected by OME. Those with cleft palate, Down syndrome, and other craniofacial anomalies are at high risk for anatomic causes of OME and worse function of the Eustachian tube.[8] Individuals of American Indian, Alaskan, and Asian backgrounds are believed to be at greater risk,[9] as are children with adenoid hyperplasia. In addition, children with existing hearing loss will be affected more dramatically by the secondary conductive hearing loss that occurs with OME.

Although rare, OME also occurs in adults. This usually happens after patients develop a severe upper respiratory infection such as sinusitis, severe allergies, or rapid change in air

pressure (after a plane flight or a scuba dive. The incidence of prolonged OME in adults is not known, but it is much less common than in children.[10]

Symptoms of Otitis Media with Effusion

OME can be associated with discomfort and a feeling of fullness in the ear. Patients with OME are also prone to episodes of acute otitis media (AOM). Temporary hearing loss is common among OME patients. This hearing loss is often mild and transient (i.e., worsened or with hearing threshold elevated by about 10 decibels [dB]), but in some cases moderate or severe prolonged hearing loss can occur.[11]

Because hearing loss in young children may delay or permanently change their communication skills and may lead to behavioral and educational difficulties,[12] clinicians and others are concerned about the possible role of OME on these outcomes. Additionally, those with chronic Eustachian tube dysfunction and OME are at risk for structural damage of the tympanic membrane.[13]

Diagnosis of Otitis Media with Effusion

Diagnostically, the core feature of OME is middle ear effusion (MEE)—i.e., fluid behind the eardrum in the middle ear space. This condition can be difficult to identify visually; however, bubbles seen behind the tympanic membrane is an almost certain indicator (i.e., pathognomonic) of the presence of MEE.[12] Tympanocentesis (use of a needle to puncture the tympanic membrane to allow for fluid drainage) remains the gold standard for diagnosing MEE and OME.

Taking a careful history is important to identify risk factors for developing OME. For example, eliciting a history of recent upper respiratory infection, allergy, subjective hearing loss or imbalance, speech and language delay, and a history of cleft palate or Down syndrome is critical.

OME is distinguished from AOM by the lack of acute symptoms or signs of inflammation. Therefore, OME should *not* have purulent fluid or redness on examination of the ear. Another distinguishing feature between AOM and OME is the appearance of the tympanic membrane, which bulges with AOM and is typically retracted or neutral with OME. Although AOM also presents with fluid behind the tympanic membrane, it is defined as also including an acute onset and symptoms and signs of middle-ear inflammation.[12]

A variety of examination techniques are used to assist with identification of OME. The most studied diagnostic method is pneumatic otoscopy; in this step, clinicians blow air through the otoscope, causing movement of the tympanic membrane. The decreased movement when fluid is present behind the TM can be identified through comparison to normal membrane movement. This is an accurate way to diagnose MEE by trained examiners.[2]

Tympanometry is a diagnostic tool that indirectly measures middle ear pressure and tympanic membrane mobility. A “flat” tympanogram (Type B tympanogram) is consistent with OME.

OME often has a corresponding conductive hearing loss on pure-tone audiometry. Hearing is generally measured across the speech range, and for young children normal hearing is considered to be no worse than 15 dB (which is the measure of loudness needed to respond to a sound).[14] In contrast, the average hearing levels for ears with OME often measure at 25 dB, with about 20 percent exceeding 35 dB.[1] Also, although usually not necessary to make a diagnosis, MEE can be demonstrated on imaging studies, such as a computed tomography (CT) scan of the temporal bone.

Treatments for Otitis Media With Effusion

Given the natural history of OME, particularly in relation to the high instance of spontaneous resolution, clinical decisions are complicated. Despite recent practice guidelines and systematic reviews,[8, 12, 15-21] the comparative benefits and harms of treatments and treatment strategies for OME are uncertain.

Surgical procedures, medications, and alternative treatments have all been used as treatments for OME. Table 1 lists the various surgical and nonsurgical interventions and overall strategies for treating OME. Treatments for OME generally try to improve Eustachian tube function (medical and physical treatments) or ventilate the middle ear while Eustachian tube function improves over time (surgical treatments).

Table 1. Treatments for otitis media with effusion, with presumed mechanism of action

Type Of Intervention	Treatment	Description	Presumed Mechanism of Action
Surgical	Tympanocentesis (or paracentesis)	A needle is used to aspirate fluid from the middle ear.	Initial relief of fluid may improve conductive hearing loss and may not recur.
	Myringotomy	After anesthesia, a small incision or perforation is made in the tympanic membrane.	Air enters the middle ear and pressure to equalize with atmospheric pressure. The hole in the tympanic membrane lasts for only a short time—i.e., is open from 1 to 10 days for standard procedure[22]
	Tympanostomy tube placement	After anesthesia (general anesthesia in children, can be topical anesthesia in adults) myringotomy is done in the tympanic membrane and a thin tube is inserted through the tympanic membrane.	Placement of the tube allows aeration of the middle ear, equalization of pressure in the middle ear, and drainage of fluid from the middle ear. Hearing and symptoms can improve allowing time for underlying Eustachian tube dysfunction to resolve.
	Adenoidectomy	After general anesthesia, the adenoids are excised from the posterior pharynx. The overlying tonsils can also be removed at the same time.	The Eustachian tube opens in the posterior pharynx in close proximity to the adenoids, and the potential benefit of removal is that the Eustachian tube function may improve thereby resolving OME.
Nonsurgical physical interventions	Autoinflation of the Eustachian tube	Using either a closed mouth and valsalva maneuver or blowing against pressure in a device against a closed glottis, the intraoral cavity pressure is increased.	Increased intraoral pressure above the Eustachian tube or middle ear pressure opens the Eustachian tube into the oropharynx. Each time the procedure is repeated, it allows intermittent aeration of the middle ear and can mitigate abnormal Eustachian tube function until function returns to normal.
	Hearing aids	A small sound amplifier is placed into the external ear canal.	This overcomes the conductive hearing loss associated with middle ear effusion. Since hearing deficit is one of the concerning effects of OME, improving hearing may eliminate adverse effects of OME.

Table 1. Treatments for otitis media with effusion, with presumed mechanism of action (continued)

Type Of Intervention	Treatment	Description	Presumed Mechanism of Action
Pharmacological interventions	Antibiotics and antimicrobials	Oral medications that kill or stop duplication of infectious agents such as bacteria are used.	Bacterial infections frequently precede or begin during OME episodes. Sterilization of middle ear fluid may decrease inflammation and allow more rapid resolution of Eustachian tube dysfunction.
	Nasal and oral steroids	Anti-inflammatory medications are applied either topically (through the nose) or systemically.	Decreased inflammation at the site of Eustachian tube orifice in the posterior pharynx or in the middle ear may improve function.
	Antihistamines	Antihistamines are used to dampen inflammatory response	See above for nasal or oral steroids.
	Decongestants	Either topical or systemic medications are used to decrease edema of mucous membranes.	Decreased swelling at or near Eustachian tube orifice may improve function.
Complementary and alternative therapies	Including, but not limited to dietary amendments and osteopathic manipulation	Varies by treatment.	Varies by treatment.
Other treatment strategies	Watchful waiting	Sometimes referred to as active observation, this involves delaying treatment while monitoring patient progress. This contrasts with immediately administering a treatment.	Not directly applicable.
	Variations in surgical technique and procedures	Clinicians may use different or possibly newer approaches or devices.	Same as those of the original or parent surgical intervention.

Abbreviations: OME = otitis media with effusion.

During topic refinement, we looked at each treatment in terms of uncertainty within the published literature (including gaps in the evidence), clinical importance, patient important outcomes, and relevance to the U.S. population.

Treatments and Treatment Strategies That Will Be Addressed in This Review: Rationale for Inclusion

The interventions described below belong to one of the four main types of treatments noted in Table 1—i.e., surgical interventions, nonpharmacologic physician interventions, pharmacotherapies, and complementary and alternative medicine (CAM) interventions. As explained more thoroughly in the methods chapter, we have adopted specific criteria for including or excluding types of studies for the different kinds of therapies; we briefly mention the included study types below.

Although the most recent guidelines for treating OME do not recommend the use of **myringotomy** alone,[12] more recent literature suggests that laser-assisted myringotomy may be a useful alternative to myringotomy plus tympanostomy tubes. These recent studies suggest that it may provide a treatment with fewer complications for selected subgroups of children and adults.[23-26] Because no systematic reviews have addressed the effectiveness of myringotomy alone, we searched for relevant randomized controlled trials (RCTs) and observational studies

examining myringotomy alone as a treatment strategy for OME in otherwise healthy children, special populations of children, and adults.

The harms and benefits of **tympanostomy tubes** for managing OME in children have been addressed by two recent systematic reviews identified during our topic refinement.[15, 17] They include a 2010 Cochrane review of 10 RCTs[17] of otherwise healthy children and a 2011 systematic review, commissioned by the Swedish Council on Technology Assessment in Health Care,[15] of 8 RCTs, not excluding special populations of children. For this review, we begin with these systematic reviews and searched for additional evidence.

A growing body of literature examines variations in tympanostomy tube-related surgical techniques and procedures for treating patients with OME. The 2011 Swedish systematic review[15] identified during our topic refinement considered various characteristics of tube design and surgical procedures. We searched for other relevant studies comparing tympanostomy tube materials, designs, and surgical procedures.

Adenoidectomy as a treatment for OME in children was also reviewed in a 2010 Cochrane Collaboration systematic review that we identified during our topic refinement.[18] The review included seven RCTs comparing adenoidectomy (with or without tympanostomy tubes) and nonsurgical management or tympanostomy tubes only; studies involved children up to 18 years of age with followup of 6 months or longer, and study populations were not limited to otherwise healthy children. Our search strategy assumed that this Cochrane 2010 review[18] identified all relevant RCT studies relating to both special populations and otherwise healthy children in the literature at the time of the review. We searched for additional evidence.

The technique of **autoinflation** has been used as a therapy for OME. The goal of autoinflation is to use either a Valsalva maneuver or external device to equalize middle ear and oropharyngeal pressure, transiently opening the Eustachian tube. A 2006 Cochrane Collaboration study identified during our topic refinement included six RCTs examining the use of autoinflation versus no treatment for hearing loss associated with OME.[19] Studies included children, adults, and special populations. We began with this review and searched for additional evidence.

The benefits and harms of **oral and topical nasal steroids** in treating children with OME and hearing loss was the focus of a 2006 Cochrane Collaboration review identified during our topic refinement.[20] The review was limited to RCTs of either steroid use alone or in combination with another agent such as antibiotics; it included special populations of children of interest to our current review. Current guidelines developed by both the United Kingdom's National Collaborating Centre for Women's and Children's Health (2008)[8] and the American Academy of Pediatrics (2004)[12] recommend against using oral or topical nasal steroids in treating children with OME. For the purposes of identifying the relevant literature for this review, we assume that the Cochrane Collaboration review identified all relevant RCTs as of the time of their review. We searched the published literature for new trials on treating OME with either oral or topical nasal steroids in children and observational studies from any time. In consultation with our Technical Expert Panel (TEP), we concluded that it would be useful to integrate newly identified studies with those previously identified through the Cochrane Collaboration review, because the newly integrated studies may result in conclusions different from those of the earlier review.[20] We conducted a completely new search to identify studies pertaining to adults, because we did not find an existing review focusing on this population.

Very little literature addresses **complementary and alternative medicine (CAM) interventions** to treat patients with OME. The book *Evidence-Based Otitis Media*[27] lists

treatments and supportive studies for at least two CAM approaches: physical manipulation and restricted diets. Based on the recommendations of our TEP, in the current review we included only RCTs of CAM interventions.

Watchful waiting, or active observation as it has more recently been called, is the process of regular review and followup of the child, including assessments of hearing, development, and educational progress. We examine this as a treatment strategy, distinct from "no treatment." Watchful waiting has not been the focus of a systematic review, although it has been a comparator in RCTs in systematic reviews focusing on other interventions. Current clinical practice guidelines recommend that watchful waiting be employed for 3 months and possibly longer, prior to initiating treatment in otherwise healthy children.[8, 12]

We considered whether to include studies reporting outcomes by ears, rather than by subjects. Omitting studies reporting results by ears is reasonable and appropriate when (1) the treatment involved is systemic or (2) outcomes are measures of the patient's overall function, such as academic achievement, speech production, language development, or quality of life. For ear-specific treatments or outcomes such as hearing thresholds or presence of fluid, ear-specific reports were included in this current review.

Treatments That Will Not Be Addressed in This Review: Rationale for Exclusion

Hearing aids are not used as a treatment option for OME in the United States. According to a 2008 National Collaborating Centre for Women's and Children's Health of the National Institute for Health and Clinical Excellence (NICE) guideline,[8] no high-quality comparative studies have evaluated the effectiveness of hearing aids to other interventions for treating OME. Furthermore, we did not find any comparative studies on hearing aids during topic refinement, and our Key Informants did not consider hearing aids of clinical relevance in the context of OME. Hearing aids, therefore, will not be included in the current review.

Using **antihistamines and decongestants** for treating children with OME has been extensively studied in primary RCTs and summarized in recent systematic reviews[21, 28] and clinical practice guidelines.[8, 12] A Cochrane review of OME for use in children identified 16 RCTs that included more than 1,800 subjects.[21] High-quality evidence on multiple short- and long-term outcomes repeatedly and unequivocally demonstrated no benefit for use of these medications over placebo for treating OME. Additionally, the reviewed studies found evidence of increased side effects and harms with the use of these medications. We see no reason to believe that these findings will change with future advances in the medication class or causes of OME. We have, therefore, decided to exclude antihistamines and decongestants from the current review as a treatment that is definitively not effective and likely harmful.

Antimicrobials including antibiotics currently are not commonly used in the United States to treat OME and are not recommended in current U.S. guidelines.[12] Some conflicting evidence exists regarding the effectiveness and utility of antibiotics for treating patients with OME.[8, 12, 28] An upcoming Cochrane Collaboration review on the use of antibiotics for the treatment of OME in children was started in 2011 and is well under way.[29] We will not duplicate their efforts and have excluded antibiotics from the current comparative review.

Scope and Key Questions

Scope

The populations, interventions, comparators, outcomes, timeframes, and settings (PICOTS) for this review are summarized in Table 2. We have indicated above which treatments we will not cover because of clear evidence of no net benefits or irrelevance to the U.S. patient population.

Table 2. Core PICOTS for this review

Population	All individuals with OME. This includes younger and older children, adolescents, adults; individuals from different racial or ethnic backgrounds; and special populations of any age including individuals with craniofacial abnormalities (e.g., cleft palate), Down syndrome, existing hearing loss, delays in speech and language, or a history of acute otitis media or OME.
Interventions	Surgical interventions: tympanostomy tubes (also referred to as PE tubes), myringotomy, and adenoidectomy, alone and in combination. Nonpharmacological and nonsurgical treatments or treatment strategies: autoinflation of the Eustachian tube. Complementary and alternative medicine interventions. Pharmacological treatments: oral or topical nasal steroids. Other treatment strategies: watchful waiting.
Comparators	Different combinations of the above interventions and strategies, including: head-to-head comparisons of one or more treatments, treatment strategies (e.g., watchful waiting vs. early treatment), or surgical procedures and techniques (e.g., one type of tympanostomy tube or procedure vs. another). Steroids only may be compared to placebo.
Outcomes	Clinical outcomes: changes in middle ear fluid, episodes of AOM, hearing thresholds, vestibular function (i.e., balance and coordination). Health care utilization: number of office visits, number of surgeries, and medication use. Functional and quality-of-life outcomes: hearing, auditory processing, speech and language, development, academic achievement, attention and behavior, quality of life, and parental satisfaction with care. Harms: all reported harms for each treatment option.
Timeframes	Short-term studies looking at outcomes from 0 to 3 months after intervention. Longer term studies looking at outcomes past 3 months and into adolescence or adulthood.
Settings	Primary care offices where the patient is seen by a pediatrician, family physician, or nurse practitioner; subspecialist physician offices where the patient is seen by an otolaryngologist; surgical settings within a hospital or outpatient clinic; emergency departments; and craniofacial treatment centers.

Abbreviations: AOM = acute otitis media; OME = otitis media with effusion; PE = pressure equalization.

We conducted this review (nominated by an adult patient) because of the continuing uncertainty about efficacy, effectiveness, and particularly comparative effectiveness, as well as harms, for the included therapies. This uncertainty leaves clinicians, patients, and families (e.g., parents of younger children) facing considerable dilemmas about choosing appropriate interventions, given patient characteristics and preferences. OME is a common condition and more up-to-date and comprehensive comparative information will be helpful to many stakeholder groups in making decisions about when and how to treat this condition. We also were mindful of the need to provide this information for populations not otherwise included in past reviews such as adults and children with special conditions such as Down syndrome, cleft palate, or existing hearing loss.

Thus, we aimed to provide useful information for clinical decisionmaking and policymaking. Of particular concern, as reflected in our KQs, were issues such as weighing benefits and harms for patients, appropriate interventions for particular population subgroups, and considering the applicability of evidence to primary versus specialty practice.

Key Questions

We addressed five key questions (KQs) in this comparative effectiveness review.

KQ 1: What is the comparative effectiveness of the following treatment options (active treatments and watchful waiting) in affecting clinical outcomes or health care utilization in patients with OME?

Treatment options include:

a. Tympanostomy tubes,
b. Adenoidectomy with or without myringotomy,
c. Myringotomy,
d. Oral or topical nasal steroids,
e. Autoinflation,
f. Complementary and alternative medical procedures,
g. Watchful waiting,
h. Variations in surgical technique or procedure.

Clinical outcomes include changes in:

a. OME signs (middle ear fluid) and symptoms (fullness in ear, difficulty in hearing), objective hearing thresholds,
b. Episodes of acute otitis media, and
c. Vestibular function such as balance and coordination.

KQ 2: What is the comparative effectiveness of the different treatment options listed in KQ 1 (active treatments, watchful waiting, and variations in surgical procedures) in improving functional and health-related quality of life outcomes in patients with OME? These outcomes include:

a. Hearing,
b. Speech and language development,
c. Auditory processing,
d. Academic achievement,
e. Attention and behavioral outcomes,
f. Health-related quality of life, and
g. Patient and parent satisfaction with care.

KQ 3: What are the harms or tolerability among the different treatment options?

KQ 4: What are the comparative benefits and harms of treatment options in subgroups of patients with OME? Subgroups include:

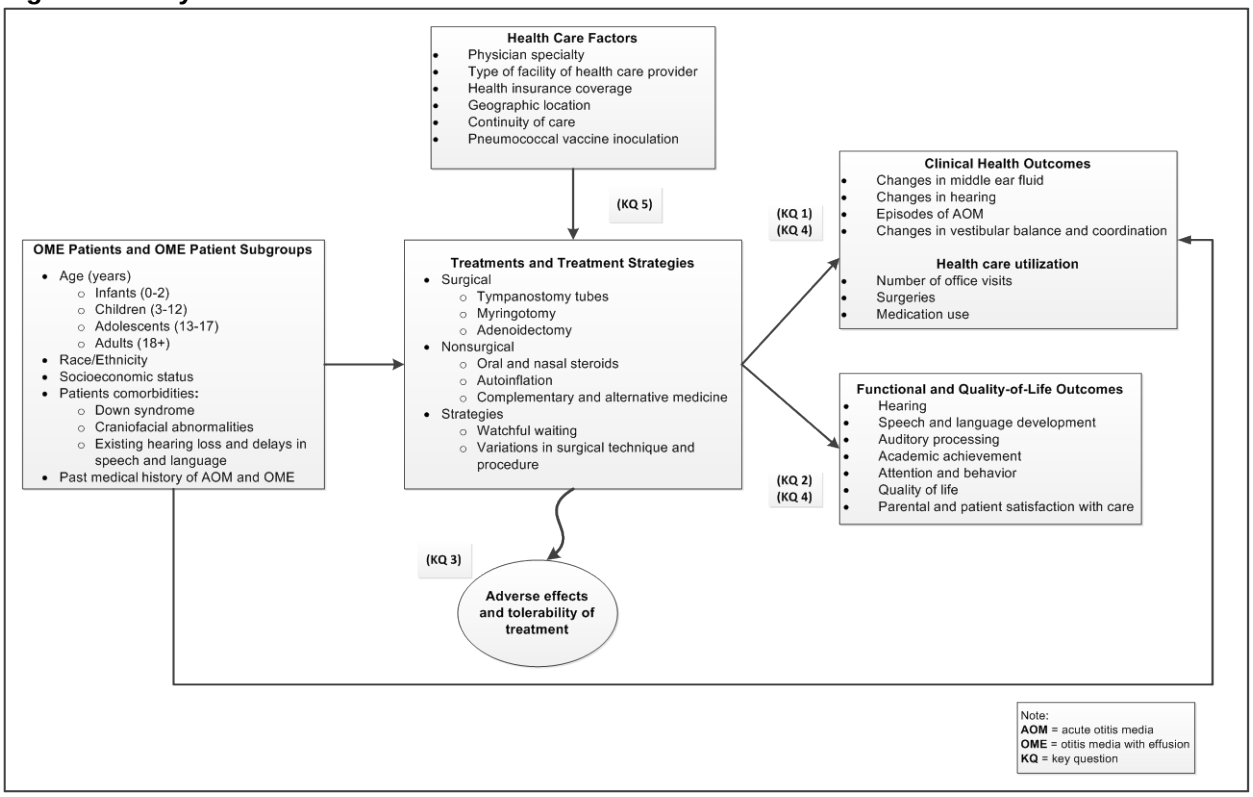
a. Patients of different age groups,
b. Patients of different racial or ethnic backgrounds,
c. Patients in different socioeconomic status groups,
d. Patients with comorbidities such as craniofacial abnormalities (e.g., cleft palate), Down syndrome, and existing speech, language and hearing problems, and
e. Patients with a medical history of AOM or OME (with and without clinical hearing loss or other problems).

KQ 5: Is the comparative effectiveness of treatment options affected by any of the following factors:

a. Health insurance coverage,
b. Physician specialty,
c. Type of facility of the treatment provider,
d. Geographic location,
e. Continuity of care, or
f. Prior inoculation with the pneumococcal vaccine?

Figure 1 gives the analytic framework for this review. The populations of interest are in the box to the far left; the interventions appear in the middle; and the two sets of outcomes (for KQ 1 and KQ 2 on benefits, and also KQ 4 on important subgroups) appear on the far right. KQ 3 concerns harm (various types of adverse events). Finally, KQ 5 relates to a set of health care delivery or clinical factors (pneumococcal vaccination) that may influence choices of treatments or their clinical and quality-of-life outcomes.

Figure 1. Analytic framework for review of treatments of otitis media with effusion



Organization of This Report

In the remainder of this report, the second chapter documents our methods, and the third chapter presents our key findings and data synthesis for all five key questions. Chapter 4 discusses findings in the light of ongoing debate and what is already known about therapy for patients with OME, discusses the limitations of the evidence base and this review, identifies gaps in the evidence, and suggests a future research agenda to fill those gaps.

The main report has several appendices, as follows: A, search strategies; B, list of studies excluded at full-text review with reasons for exclusion; C, evidence tables; D, abstract and full text forms; E, risk of bias tables; F, strength of evidence tables; G, glossary; and H is an acronyms list.

Methods

The Evidence-based Practice Center (EPC) conducted this review using the research methods described in the Agency for Healthcare Research and Quality (AHRQ) Methods Guide for Effectiveness and Comparative Effectiveness Reviews.[30] Further, we used the PRISMA Statement as a guide to ensure transparent reporting.[31]

Topic Refinement and Protocol Review

The EPC developed this topic and key questions through a public process. The topic was nominated through an online public forum and subsequently developed and refined by a team at the EPC with input from Key Informants in the field. AHRQ posted key questions for public comment (11/17/2011). We incorporated public comments and guidance from a Technical Expert Panel (TEP) into the final research protocol, which was also posted on the AHRQ Web site (3/20/2012).

Literature Search Strategy

Search Strategy

During topic refinement, the EPC identified five recently published systematic reviews with results on comparisons of interest for otitis media with effusion (OME) that were conducted either by the Cochrane Collaboration or commissioned by a national governmental agency. The Cochrane Collaboration conducted four of these;[17-20] the Swedish Council on Technology Assessment in Health Care was commissioned the fifth.[15] These reviews covered the following OME-related treatment topics: autoinflation, nasal steroids, tympanostomy tubes, and adenoidectomy.

To avoid repeating or duplicating the work of these other systematic review teams, we limited our search, review, and analysis for each of our key questions (KQs) to evidence that these systematic reviews included plus evidence from other reports that these recent reviews would not have considered. These newer elements of our review include observational studies done at any time such as nonrandomized trials, newer trials published since the last search dates in those reviews, and studies focusing on populations excluded from the reviews, such as adults with OME or children with Down syndrome or cleft palate, who may be differently affected by OME.

We conducted focused searches of MEDLINE® (via PubMed), EMBASE, CINAHL (nursing and allied health database) and the Cochrane library. An experienced research librarian used a predefined list of search terms and medical subject headings (MeSH). The librarian completed the first search on 1/8/2012 and will conduct an update search during peer review. We limited searches to studies published in English, given limited resources. The complete search strategies, including specific limitations used for each database, are presented in Appendix A.

We searched unpublished and grey literature relevant to the review topic. Methods for identifying grey literature included a review of trial registries, specifically ClinicalTrials.gov, Health Services Research Projects in Progress (http://www.nlm.nih.gov/hsrproj/), and the European Union Clinical Trials Register (https://www.clinicaltrialsregister.eu/). Further, AHRQ requested Scientific Information Packets (SIPs) from the developers and distributors of the interventions identified in the literature review. SIPs allow an opportunity for the intervention

developers and distributors to provide the EPC with both published and unpublished data that they believe should be considered for the review. We included unpublished studies that met all inclusion criteria and contained enough information on their research methods to permit us to make a standard risk-of-bias assessment of individual studies.

Lastly, we searched reference lists of review articles that are pertinent but did not meet inclusion criteria for studies that we should consider for inclusion in this review.

Inclusion and Exclusion Criteria

Table 3 outlines the Population, Intervention, Comparators, Outcomes, Timing, and Settings (PICOTS) that define the major inclusion criteria for studies in this review. In the following sections we provide additional detail related to each of these domains as needed.

Table 3. Inclusion and exclusion criteria for studies of Otitis Media with Effusion

Domain	Description
Population	• All individuals with OME. Subpopulations include infants; adults; individuals from different racial/ethnic backgrounds; and special populations of any age including individuals with craniofacial abnormalities (e.g., cleft palate), Down syndrome, existing hearing loss, delays in speech and language, or a history of AOM or OME.
Interventions	• Surgical interventions: tympanostomy tubes (also referred to as pressure equalization tubes, grommets and ventilation tubes), myringotomy (also referred to as paracenteisis), and adenoidectomy with or without myringotomy. • Pharmacological treatments: oral or topical nasal steroids. • Nonpharmacological and nonsurgical treatments or treatment strategies: watchful waiting, complementary and alternative medicine procedures, and autoinflation of the Eustachian tube.
Comparator	• Different combinations of the above interventions and strategies. These include head-to-head comparisons of one or more treatments, treatment strategies (e.g., watchful waiting vs. early treatment), or surgical procedures and techniques (e.g., one type of tympanostomy tube or procedure vs. another or different adjunct therapies to enhance the main intervention). We considered inactive controls in comparison with steroid treatment and usual care in comparison with autoinflation, based on the Cochrane Collaboration systematic review inclusion criteria. We considered head-to-head trial evidence and observational study data.
Outcomes	• Clinical outcomes: changes in middle ear fluid, episodes of AOM, hearing thresholds, vestibular function (i.e., balance and coordination). • Health care utilization: number of office visits, number of surgeries, and medication use. • Functional and quality-of-life outcomes: hearing, auditory processing, speech and language development, academic achievement, attention and behavior, quality of life, and parental satisfaction with care. • Harms: all reported harms for each treatment option.
Timing	• Shorter studies looking at outcomes 0 to less than 3 months post intervention. • Longer studies looking at outcomes past 3 months and into adolescence or adulthood.
Setting	• Studies conducted in the United States or internationally. • Interventions provided in primary care offices where the patient is seen by a pediatrician, family physician, or nurse practitioner; subspecialist physician offices where the patient is seen by an otolaryngologist; surgical settings within a hospital or outpatient clinic; emergency departments; and craniofacial treatment centers.

Abbreviations: AOM = acute otitis media; OME = otitis media with effusion.

Population

The population of interest for this review included individuals with OME, defined as a collection of fluid in the middle ear without signs or symptoms of ear infection. Patients had to have OME at the time of the intervention. We excluded studies that focused on the interventions of interest, such as tympanostomy tubes or myringotomy, but did not isolate results for

individuals with only OME, because we could not measure the results in the OME population. Most commonly, studies with a mix of participants also included individuals with acute otitis media (AOM). For the same reason, we also excluded studies that focused on subpopulations of interest, such as adults or children with craniofacial abnormalities, if all participants did not all have a diagnosis of OME.

Intervention

Interventions were limited to the surgical, pharmaceutical, and nonpharmaceutical listed in Table 2. Interventions could include a combination of these interventions, such as adenoidectomy and tympanostomy tubes. Interventions could also include adjunct therapy, such as topical substances to reduce the harms from tympanostomy tubes.

Comparators

All studies included in this review had to have at least two arms, and all participants had to have been diagnosed with OME. Acceptable comparisons included one of the other treatment comparisons included in the review, except that for steroid treatment, we included placebo or nonintervention controls because these were the only comparison studies available. Autoinflation treatment was considered in comparison without autoinflation with the addition of usual care treatments, provided they were administered equally in both arms.

Studies that included adjunct therapies that were not the focus of the review, such as antibiotic treatment, were included if those therapeutic modalities were provided similarly to all study arms.

Outcomes

Study outcomes were categorized as clinical (KQ 1), functional (KQ 2), and harms (KQ 3), corresponding to our (KQs). Clinical outcomes were grouped as OME signs and symptoms, objective hearing, AOM, vestibular function such as balance and coordination, and use of health care services. Functional outcomes were grouped as speech, language, and cognitive development, behavior, quality-of-life, and satisfaction with care. Potential harms differ across interventions (i.e., surgical, pharmaceutical, device)

Timing

We included studies reporting outcomes of less than 3 months and 3 months or longer including some with only end-of-intervention results.

Setting

We did not exclude studies based on geography or the setting of service provision.

Study Designs

Table 4 describes the study design inclusion criteria developed for this report.

Table 4. Study inclusion criteria for review of otitis media with effusion

Category	Criteria for Inclusion
Study design	Meta-analyses, systematic reviews conducted by the Cochrane Collaboration or commissioned by a national governmental agency that were identified during topic refinement, RCTs, and nonrandomized controlled trials, prospective and retrospective cohort studies and case-control studies not included in one of these 5 systematic reviews.
Study duration	Unlimited
By ear or by subject studies	Studies could separate groups by subject or by ear. For studies by ear to be considered RCTs, they needed to randomize by ear. Studies that analyzed results by ear and created groups by distinguishing between left ear and right ear are considered nonrandomized controlled trials.
Sample size	Unlimited
Study location	Unlimited
Time of publication	Because some of the treatment options of interest have been comprehensively addressed in recent Cochrane Collaboration or national government-commissioned systematic reviews, we searched only for new literature and observational studies when a treatment had been addressed in a review from one of these two types of sources. The following summarizes our search strategy for each included treatment option and population of interest. We searched from 1948 forward for: • All treatments not addressed in one of the identified systematic reviews. This included studies comparing different surgical techniques or different types of the same intervention (i.e., tympanostomy tubes, myringotomy). • Nonrandomized and observational studies across treatment options. • Studies concerning adults and subpopulations of interest (particularly children with comorbidities such as Down syndrome and craniofacial abnormalities), across treatment options. • RCTs of complementary and alternative medicine • RCTs of treatments covered in systematic reviews - Tympanostomy tubes vs. nonsurgical interventions In relation to otherwise healthy children, who would not be considered as members of subpopulations of interest because of comorbidities, we included evidence from two recent systematic reviews relevant to our KQs and searched all new RCT literature published one year prior to the last search, April 2006 forward. - Adenoidectomy with or without myringotomy In relation to otherwise healthy children, who would not be considered as members of subpopulations of interest because of comorbidities, we included evidence from two recent systematic reviews relevant to our KQs and searched all new RCT literature published one year prior to the last search, March 2008 forward. - Oral and topical nasal steroids In relation to otherwise healthy children, who would not be considered as members of subpopulations of interest because of comorbidities, we included evidence from two recent systematic reviews relevant to our KQs and searched all new RCT literature published one year prior to the last search, May 2005 forward. - Autoinflation In relation to otherwise healthy children, who would not be considered as members of subpopulations of interest because of comorbidities, we included evidence from two recent systematic reviews relevant to our KQs and searched all new RCT literature published one year prior to the last search, August 2005 forward.
Language of publication	Given the volume of literature on this topic, we limited our search to publications in the English language.

*Search to be updated when report is out for peer review.

Abbreviations: KQs = key questions; RCT = randomized controlled trial.

Study Selection

Two trained members of the research team independently reviewed all titles and abstracts produced by the searches to determine study eligibility against predefined inclusion and exclusion criteria. Studies marked for possible inclusion by either reviewer underwent a full-text review. Each full-text article was again independently reviewed by two trained members of the team to determine if it met inclusion criteria. If both reviewers agreed that a study did not meet the eligibility criteria, it was excluded; each reviewer recorded the primary reason for exclusion. If the reviewers disagreed, they resolved conflicts by discussion and consensus or by consulting a third member of the review team. The full-text review form reviewers used is reproduced in Appendix B.

The project coordinator tracked results of the abstract and full-text reviews in an EndNote database. Appendix B contains a complete list of studies excluded during the full-text review, denoted by their primary reason for exclusion.

We screened unpublished studies identified through grey literature search and review of SIPs using the same title/abstract and full-text review processes.

Data Extraction

We developed a template for evidence tables for data synthesis using the PICOTS framework. For the systematic reviews and additional studies that met inclusion criteria, we abstracted relevant information into these evidence tables using Microsoft Excel. We abstracted characteristics of study populations, interventions, comparators, settings, study designs, methods, and results. Data from studies included in the systematic reviews were abstracted as they were presented in the review, although we did refer to the original article to obtain additional information for clarification purposes. We also reviewed the original studies included in the systematic reviews to determine if additional data concerning subgroup analyses were contained in any of the studies and not reported in the overall systematic review results. One trained reviewer initially abstracted the relevant data from each included article and a second member of the team reviewed each data abstraction against the original article for completeness and accuracy.

Risk of Bias Assessment

For each included systematic review and study, we assessed the potential for selection bias, performance bias, attrition bias, detection bias, and outcome reporting bias using instruments that have been successfully used previously by our EPC (Appendix tables E-1 through E-5). The risk of bias assessment was conducted using two tools, one appropriate for trials or observational studies (some questions concerning trial study design would be considered not applicable) and one appropriate for systematic reviews. Both tools were based on instruments that had been used successfully in earlier reviews conducted by our EPC. We did not reevaluate the risk of bias of the individual studies included in the five systematic reviews and relied on the original authors' assessments. Two independent reviewers rated the risk of bias for each study. Disagreements between the two reviewers were resolved by discussion and consensus or by consulting a third member of the team.

Results of this assessment are summarized in a rating of low, medium, or high risk of bias. In general, a study with a low risk of bias has a strong design (adequate randomization and allocation concealment if a trial and controls for concurrent treatments), measures outcomes

appropriately including whether there was blinding of the patient and provider (if possible) and outcome assessor, uses appropriate statistical and analytical methods, reports low attrition, and reports methods and outcomes clearly and precisely. Studies with a medium risk of bias are those that do not meet all criteria required for low risk of bias but do not have flaws that are likely to cause major bias. Studies with a high risk of bias include those with at least one major issue that has the potential to cause significant bias and thus might invalidate the results. Examples of flaws leading to a high risk of bias rating include different application of inclusion/exclusion criteria between groups, substantial differences in groups at baseline, high overall attrition, or differential attrition across study conditions, lack of control for concurrent treatment or among cohort studies, lack of control for critical potential confounding, either through design or statistical analyses. A high risk of bias rating was assigned to studies in which the critical information needed to make that assessment was not reported or was unclear. To maintain a focus on interpretable evidence, we opted to not include studies with a high risk in the synthesis of findings in the results chapter of this review. We list each study rated as high risk of bias and the main reason we gave it that rating in Appendix E.

Data Synthesis

Across all included studies, the populations, interventions, and outcome measures in the additional data were heterogeneous and did not lend themselves to a pooled analysis. They also did not lend themselves to updating the meta-analyses from the five earlier systematic reviews. Thus, we did all analyses qualitatively. Evidence used in the synthesis included the results from the earlier meta-analyses and additional data from individual studies as presented in the systematic reviews, and data from the articles included from our own searches.

Strength of the Body of Evidence

In the key points section we present the strength of evidence for each comparison and overarching outcome (e.g., OME signs and symptoms, hearing,) as specified for each KQ. We graded the strength of evidence based on the guidance established for the AHRQ Effective Health Care Program EPCs conducting comparative effectiveness reviews, as detailed in the paper by Owens and colleagues.[30] The EPC approach incorporates four key domains: risk of bias, consistency, directness, and precision of the evidence.

- Risk of bias is determined according to the "degree to which the included studies for a given outcome or comparison have a high likelihood of adequate protection against bias." It is graded as high, medium, or low.
- Consistency is the "degree to which reported effect sizes from included studies appear to have the same direction of effect." Each body of evidence is graded as consistent or inconsistent. Consistency cannot be assessed when a body of evidence has only a single study (unknown or not applicable). When studies included in a body of evidence include both consistent and inconsistent findings, the presence of one or more consistent findings will result in a "consistent" grade for the outcome of interest.
- Directness is determined based on "whether the evidence links the interventions directly to health outcomes." It is graded direct or indirect. In this review, most of the included measures are direct. When a body of evidence includes both indirect and direct measures, the presence of one or more direct measures will result in a "direct" grade.

- Lastly, precision is determined according to "the degree of certainty surrounding an effect estimate." for each outcome separately. "Precise" indicates a clinically useful conclusion that is statistically significant, and "imprecise" indicates that no conclusion can be drawn as to whether either treatment is superior or whether the treatments are equivalent.

The overall grades for strength of evidence, based on the scores for the above domains, are described in Table 5. Grades reflect the strength of the body of evidence to answer the KQs on the comparative effectiveness, efficacy, and harms of the interventions in this review.

Table 5. Definitions of the grades of overall strength of evidence

Grade	Definition
High	High confidence that the evidence reflects the true effect. Further research is very unlikely to change our confidence in the estimate of effect.
Moderate	Moderate confidence that the evidence reflects the true effect. Further research may change our confidence in the estimate of the effect and may change the estimate.
Low	Low confidence that the evidence reflects the true effect. Further research is likely to change our confidence in the estimate of the effect and is likely to change the estimate.
Insufficient	Evidence either is unavailable or does not permit estimation of an effect.

Source: Owens et al., 2010[30]

Two reviewers assessed each domain independently and also assigned an overall grade for each key outcome listed in the framework; they resolved any conflicts through consensus discussion. If they did not reach consensus, the team brought in a third party to settle the conflict.

Applicability

We assessed the applicability both of individual studies and of the body of evidence. For individual studies, we examined factors that may limit applicability based on the PICOTS structure. Examples of characteristics examined include:

Population
- Narrow eligibility criteria, or exclusion of patients with comorbidities;
- Large differences between demographics of the study population and community patients.

Intervention
- Intensity and delivery of interventions that may not be feasible for routine use;
- Highly selected intervention team or level of training/proficiency not widely available.

Comparators
- Comparison group does not represent an available alternative treatment.

Such factors may be associated with heterogeneity of treatment effect and may lessen our ability to generalize the effectiveness of an intervention to use in everyday practice. We abstracted key characteristics of applicability into evidence tables.

During data synthesis, we assessed the applicability of the body of evidence using the abstracted characteristics. KQ 4 includes an analysis of intervention efficacy in population subgroups.

Peer Review and Public Commentary

Experts in OME, specifically clinicians and researchers specializing in ear, nose and throat treatment, pediatrics, and audiology, and evidence-based interventions, will be invited to provide external peer review of the draft comparative effectiveness review. AHRQ and an associate editor also provided comments. The draft report will be posted on the AHRQ Web site for 4 weeks to elicit public comment. We will respond to all reviewer comments and note any resulting revisions to the text in the "Disposition of Comments Report." This disposition report will be made available 3 months after the final CER is the posted on AHRQ Web site.

Results

Introduction

This chapter first presents the results of our literature searches. We then take up the findings of our analyses for each key question (KQ) in turn; we address the following as relevant to the KQ, in this order:

- Surgical procedures, specifically:
  - tympanostomy tubes, comparisons of different types or insertion approaches
  - tympanostomy tubes versus myringotomy or nonsurgical interventions (delayed treatment or watchful waiting),
  - tympanostomy tubes plus adenoidectomy versus myringotomy plus adenoidectomy or adenoidectomy alone
  - myringotomy versus myringotomy, comparison of different approaches, various combinations of myringotomy plus adenoidectomy; and
  - adenoidectomy versus nonsurgical interventions or tympanostomy tubes
- Nonpharmacological interventions, specifically, autoinflation
- Other treatment strategies, specifically delayed treatment or watchful waiting are presented in comparison to other treatment approaches above.

We did not find any randomized controlled trial (RCT) evidence concerning complementary or alternative medicine (CAM) treatments or procedures and therefore, this intervention will not be discussed further.

A description of all included studies, for each treatment comparison, is presented at the beginning of the results for KQ 1. Because KQ 1 includes all studies, a description of studies is not repeated for other KQs. We then present key points with grades for strength of evidence for major comparisons and outcomes; these are followed by text and tables providing a more detailed synthesis of the included studies. When no studies reported on categories of outcomes, we note this in key points and do not repeat that point in detailed synthesis.

All results new to this systematic review are based on qualitative synthesis. Because of the heterogeneity of populations, interventions, or outcomes in the included studies, we could not do any new quantitative (pooling) analyses. We present results from meta-analyses that were conducted in the five earlier systematic reviews that are included as part of our evidence.

We included in these analyses only studies that we had rated as low or medium risk of bias. All studies are medium risk of bias unless otherwise specified as low. In summary tables that describe included studies, we specify not only study type (e.g., RCT, nonrandomized trial or observational study) but also whether, for RCTs, the randomization was done by ear or by participant. Studies rated high risk of bias are listed in Appendix E together with the principal reason(s) for that rating. Evidence tables for included studies can be found in Appendix C and include the risk of bias assessments for each of the included studies and systematic reviews. Detailed strength of evidence tables are presented in Appendix F. The final strength of evidence grades for the most critical findings are presented in this chapter.

A description of procedures for measuring hearing, language and quality of life measures may be found in Table 6.

Table 6. Description of procedures/measures of hearing, language and quality of life related to OME

Method of measurement and example indices	Description	Range/Meaning of Possible Scores	Improvement Indicated by
Hearing Measures: Air-Bone Gap (ABG)	A method of diagnosing conductive hearing loss. It is the difference in audiometric hearing thresholds using bone conduction and air conduction.	The degree of conductive hearing loss is represented by difference in audiometric hearing thresholds using bone conduction in which sound transmission bypasses the middle ear and air conduction. Greater ABGs indicate grater hearing loss.	Reductions in ABGs
Hearing Measures: Pure-tone audiometry (PTA)	PTA is a behavioral test used to measure hearing sensitivity. Pure-tone thresholds (PTTs) or hearing levels (HLs) indicate the softest sound audible to an individual at least 50% of the time. Results are often averaged over different frequency levels. A modified form is sweep audiometry.	Normal hearing is age dependent: 15dB for young children, 20 dB for children through early adolescence; and 25 dB for older adolescents and adults.	Reduction in PTA HLs
Hearing Measures: Sweep audiometry	A modified form of pure tone audiometry.	Same interpretation as PTA	Same as PTA
Hearing Measures: Speech Recognition Threshold (SRT)	The speech recognition threshold is the softest level at which speech is understood.	Scores are given in dBs and have the same meaning as pure tone hearing levels	Reduction in SRTs
Speech and Language: Receptive language Reynell Developmental Language Scales - Verbal Comprehension Scale - Expressive Language Scale	Receptive language measures how one understands language	Usually provided as a standard score that has been normed on a representative sample	Increases in standard scores
Speech and Language: Expressive language, Reynell Developmental Language Scales	Expressive language measures how one produces oral language	Usually provided as a standard score that has been normed on a representative sample	Increases in standard scores
Quality of Life: Otitis Media 6 (OM-6)	Parent reported scale measuring effects of OME on quality of life.	1-7, higher scores associated with poorer quality of life	Decreases

Results of Literature Searches

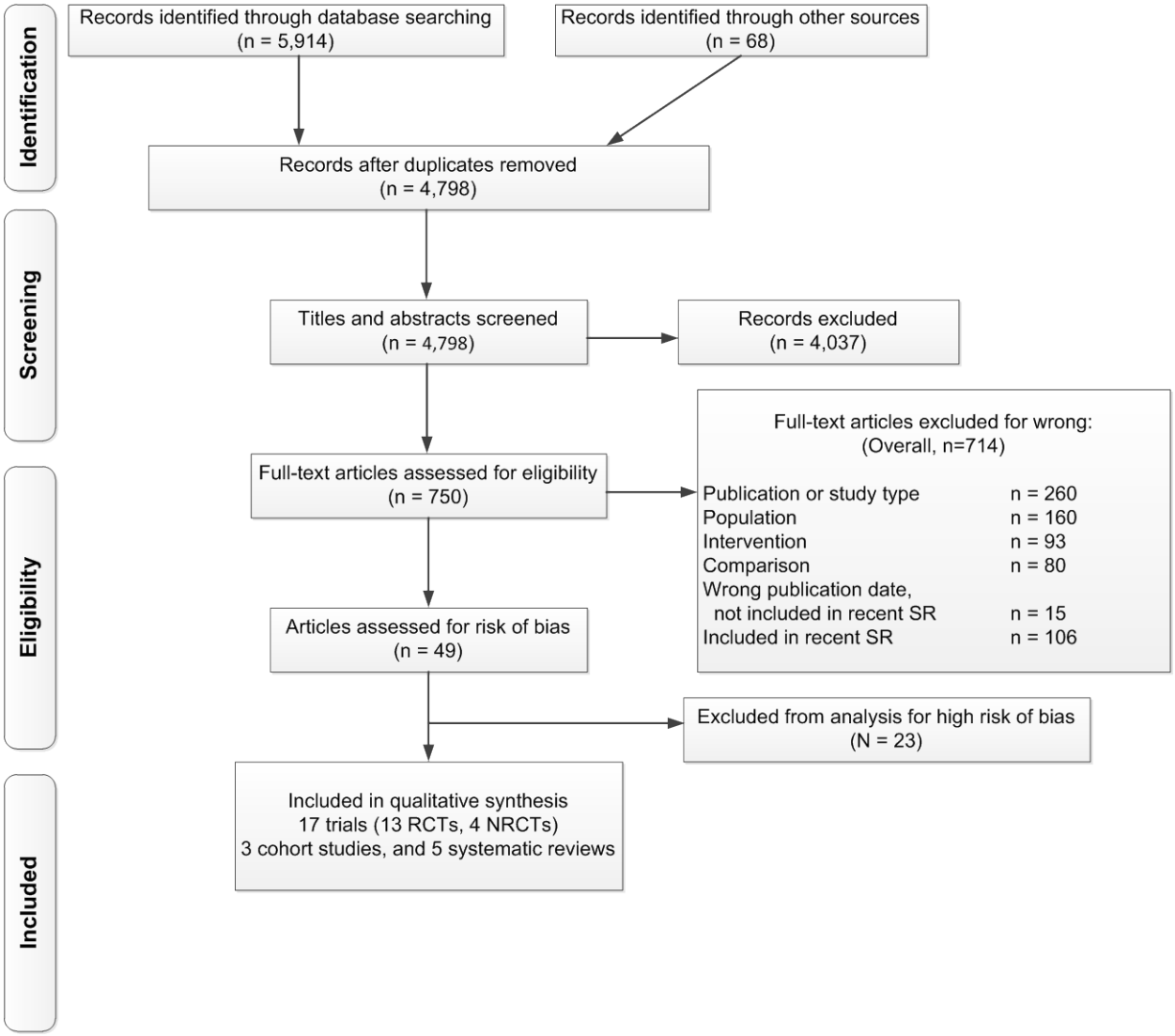
Figure 2 presents our literature search results. Literature searches through 2/28/2012 for the current report identified 4798 unduplicated citations. Appendix A provides a list of all search terms used and the results of each literature search.

After applying our eligibility and exclusion criteria to titles and abstracts of all identified citations, we selected 750 citations for full text review. We reapplied our inclusion criteria and excluded 701 of these articles from further review before risk of bias assessment. Appendix B provides a list of excluded studies and reasons for exclusion at the full-text stage.

Of the 49 publications included after full-text review (44 articles and 5 systematic reviews), we dropped 23 articles from further analysis because of their high risk of bias. Thus, we included

a total of 20 trials reported in 21 articles and 5 systematic reviews for qualitative synthesis. Evidence tables for these articles and systematic reviews are provided in Appendix C and risk of bias assessments can be found in Appendix D. Risk of bias assessments are also provided for the 23 high risk of bias studies in Appendix D.

Figure 2. Disposition of articles



Abbreviations: NRCT = non-randomized controlled trial; RCT = randomized controlled trial; SR = systematic review

Of the 20 studies (21 articles) included in this review, 13 were RCTs, four were non-randomized control trials, and three were retrospective cohort studies. We assessed 19 included studies as medium risk of bias and one as low risk of bias. Of the five included systematic reviews, four were limited to RCTs. We assessed four systematic reviews as low risk of bias and one as medium risk of bias (Appendix E, Table E-3 presents details of these assessment).

KQ 1: Comparative Effectiveness of Interventions: Clinical Outcomes or Health Care Utilization

Surgical Interventions: Tympanostomy Tube Comparisons

Description of Studies

The included evidence about comparisons of different types of tympanostomy tubes consisted of one recent systematic review[15] and seven additional studies[35-40] (Table 7). All studies compared groups by "ears." Of the seven additional studies we identified, four were RCTs,[35, 36, 39, 41] one was a nonrandomized controlled trial,[40] and two were observational studies.[37, 38] The Hellstrom et al.[15] systematic review included RCTs and nonrandomized control trials (i.e., systematically assigning one intervention to the left ear and the other for right ear). Tympanostomy tube comparisons included tube design (i.e., material, coatings, shape, and size) and routes or techniques for insertion. Broadly speaking, tympanostomy tubes are often categorized by expected length of time they can be expected to stay in place—chiefly, short- or long-term.

Table 7. Characteristics of studies: Tympanostomy tube comparisons

Study, Study Type, Country	Arm (N)	Diagnosis Criteria	Wait Period Between Diagnosis and Study	Inclusion/ Exclusion Criteria	Length of Follow-up	Age (Range)	Risk of Bias
Licameli et al., 2008[39] RCT by ear United States	G1: Phophoryl-choline-coated fluoroplastic Armstrong tube (N=70) G2: Uncoated fluoroplastic Armstrong tube (N=70)	NR	34 mos	Include: OME w/3-4 mos medical management	24 mos	Mean: 19 mos (8-51 mos)	Medium
McRae et al., 1989[41] RCT by ear England	G1: Shah Teflon tube + aspiration prior to placement (N=55) G2: Shah Teflon tube (no aspiration) (N=55)	Otoscopy and impedence audiometry	NR	Include: OME	24 mos	Mean: 5.7 years (2-10 years)	Medium
Oversen et al., 2000[35] RCT by person and by ear Demark	G1: TT + N-acetylcysteine instilled (N=37) G2: TT + placebo vehicle (N=38)	Otiomicrosco-pical examinations including tympanometry	3 mos	Include: OME, pressure <200mmHg Excluded: Recent antibiotics or AOM at time of surgery	39 mos	Mean: 38 mos (1-7 years)	Medium

Table 7. Characteristics studies: Tympanostomy tube comparisons (continued)

Study, Study Type, Country	Arm (N)	Diagnosis Criteria	Wait Period Between Diagnosis and Study	Inclusion/ Exclusion Criteria	Length of Follow-up	Age (Range)	Risk of Bias
Wielinga and Smyth, 1990[36] RCT by ear Ireland	G1: Goode Silicon tube (N=15) G2: Teflon Armstrong tube (N=15)	Otoscopy, PTA, tympanometry	6 mos	Include: 6 mos unsuccessful tx with standard decongestive meds; mucoid secretion aspiration, 1 patient adenoidectomy	7 years	Male mean: 7 years Female mean: 6 years	Medium
Abdullah et al., 1994[40] NRCT by ear England	G1: Trimmed high-grade silicone Shah permavent tube (N=25) G2: Polyethylene Shah tube (N=25)	NR	NR	Include: Age 3-10 years, de novo MME Exclude: History of significant AOM	29 mos	Mean: 6 years	Medium
Iwaki et al., 1998[37] Observational by ear Japan	G1: Teflon Shepard tube (N=75) G2: Silicone Goode-T tube (N=39) G3: Silicone Paparella II tube (N=106)	Conductive hearing loss >25dB air-bone gap, Type B tympanogram	6 mos	Include: 25dB air-bone conductive hearing loss, Type B tympanogram, failed politerization and conservative management	24 mos	Mean yrs: G1: 6.2 G2: 6.2 G3: 5.8 (3-12)	Medium
Slack et al., 1987[38] Observational by ear England	G1: Shepard tube (N=214) G2: Shah tube (N=70) G3: Paparella tube (N=275) G4: Goode tube (N=4) G5: Reuter Bobbin TT (N=28) G6: Other type (N=63)	NR	NR	Include: OME, <16 years of age	30 mos	NR	Medium

Table 7. Characteristics of studies: Tympanostomy tube comparisons (continued)

Study, Study Type, Country	Arm (N)	Diagnosis Criteria	Wait Period Between Diagnosis and Study	Inclusion/ Exclusion Criteria	Length of Follow-up	Age (Range)	Risk of Bias
Hellstrom et al., 2011[15] Systematic Review International	Arms differ across comparisons: 2 trials (262 participants)	Swedish guidelines for diagnosis of OME as a painless inflammation with effusion in the middle ear with impaired hearing ≥ 3 mos; no other criteria specified	Minimum of 3 mos	Include: RCTs (individual or ear), NRCTs, and cohort studies published between 1966 and 2007 of efficacy of tubes on hearing, language development, and quality of life and of complications	Various	Children or adolescents	Medium

Abbreviations: AOM = acute otitis media; dB = decibels; G = group; NR = not reported; OME = otitis media with effusion; PTA = pure tone audiometry; RCT = Randomized controlled trial; tx = treatment; TT = tympanostomy tube.

Key Points

- Single small studies compared different types of tubes, approaches to insertion, or topical prophylaxis therapies for tube retention, OME recurrence, and hearing. Evidence was insufficient.
- No studies compared time with middle ear effusion, vestibular outcomes, or health care service use as a function of type of tympanostomy tube or routes or techniques in their insertion. The strength of evidence is considered insufficient because there are no studies.

Detailed Synthesis

Table 8 presents comparisons of tympanostomy tubes in relation to clinical and health care utilization outcomes. The Hellstrom et al.[15] systematic review identified two RCTs of children with OME who had different types of tubes placed and reported on outcomes including tube retention, OME recurrence, or other ear symptoms.[15] These included a study by Hampal et al. (1991) that compared the mini-Shah tube with a standard Shah tube; the standard Shah tube had a significantly lower rate of OME recurrence.[42] Heaton et al. (1991) compared Shepard tubes (considered shorter term tubes) and Sheehy tubes (considered longer term tubes). Sheehy tubes were retained significantly longer than Shepard tubes ($p<0.001$).[43]

Table 8. Clinical outcomes: Tympanostomy tube comparisons

Study	Arm (N)	Study Duration Until Outcome Measurement	Tube Retention (% Retained Unless Otherwise Noted)	OME Recurrence	Measured Hearing
Wielinga and Smyth, 1990[36]	G1: Goode Silicon tube (N=15) G2: Teflon Armstrong tube (N=15)	NR	Mean months (range) G1: 52.5 (5-88) G2:17.5 (1-56) (p=NR)	NR	Mean: G1: 14dB G2: 11dB (p=NR)
		Year 1	G1: 93 G2: 67 (p=NS)	NR	NR
		Year 2	G1: 80 G2: 13 (p<0.05)	NR	NR
		Year 3	G1: 73 G2: 7 (p<0.05)	NR	NR
		Year 4	G1: 53 G2: 7 (p<0.05)	NR	NR
		Year 5	G1: 33 G2: 0 (p=NS)	NR	NR
		Years 6 & 7	G1: 27 G2: 0 (p=NS)	NR	NR
Abdullah et al., 1994[40]	G1: Trimmed high grade silicone Shah permavent tube (N=25) G2: Polyethylene Shah tube (N=25)	Month 12	G1: 100 G2: 56 (p=NR)	NR	NR
		Month 29	G1: 71 G2: 18 (p=NR)	G1: 6% G2: 53% (p=NR)	NR
Licameli et al., 2008[39]	G1: Phophoryl-choline-coated fluoroplastic Armstrong tube (N=70) G2: Uncoated fluoroplastic Armstrong tube (N=70)	Year 2	G1: 21 G2: 28 (p=0.84)	NR	NR
Iwaki et al., 1998[37]	G1: Teflon Shepard tube (N=75) G2: Silicone Goode-T tube (N=39) G3: Silicone Paparella II tube (N=106)	Seen at 1-3 month intervals post surgery and at 1-3 month post tube removal or extrusion	Mean months G1: 5.9 G2: 10.7 G3: 15.1 (p=NR)	NR	NR
	NR	24 months	G1: 9.3 G2: 20.5 G3: 50 (p=NR)	G1: 40% G2: 28.2% G3: 17.0% (p<0.01)	NR

Table 8. Clinical outcomes: Tympanostomy tube comparisons (continued)

Study	Arm (N)	Study Duration Until Outcome Measurement	Tube Retention (% Retained Unless Otherwise Noted)	OME Recurrence	Measured Hearing
Oversen et al., 2000[35]	G1: TT + N-acetylcysteine instilled (N=37) G2: TT + placebo vehicle (N=38)	Seen 1, 3, 6, 9, 12, 24, and 36 months post surgery	G1: mean 9 G2: mean 7 (p=0.14)	G1: 16% G2: 13% (p=NR)	NR
McRae et al., 1989[41]	G1: Shah Teflon tube + aspiration before placement (N=55) G2: Shah Teflon tube (no aspiration) (N=55)	3 months	G1: 90 G2: 92 (p=1.0)	NR	NR
		6 months	G1: 76 G2: 80 (p=0.71)	NR	NR
		12 months	G1: 47 G2: 41 (p=0.71)	NR	NR
		18 months	G1: 7.8 G2: 5.8 (p=1.0)	NR	NR
		24 months	G1: 2 G2: 0 (p=1.0)	NR	NR
Hellstrom et al., 2011[15] Systematic Review	1 study (N=116) G1: Shah tube G2: Mini-Shah	Year 1	G1: NR G2: NR P<0.001, favors G1	G1: NR G2: NR P<0.05, favors G1	
	1 study (N=146) G1: Shepard tube G2:Sheehy tube	Year 2	G1:NR G2:NR P< 0.0001, favors G2		

Abbreviations: dB = decibels; G = group; NR = not reported; NS = not statistically significant; OME = Otitis media with effusion; TT = tympanostomy tubes; vs. = versus.

Wielinga et al. demonstrated that after 1 year, silicone Goode-T tubes were more likely to be retained than Teflon Armstrong tubes.[36] Comparing two types of Shah tubes, Abdullah et al. demonstrated that at 29-month followup, silicone permavent tubes had a 71 percent retention rate as contrasted with an 18 percent rate for the polyethylene Shah tubes.[40] Similarly, in an observational study, Iwaki et al. found that Goode-T tubes were retained on average 5 months longer than Shepard Teflon tubes, but that Paparella II tubes were retained the longest (10.7 vs. 5.9 vs. 15.1 months in place, respectively).[37]

Licameli et al. compared extrusion rates in Armstrong tubes (short-term tubes), with and without phosphorylcholine-coated fluoroplastic; the groups did not differ at 2 year followup (21% vs. 28%, respectively, p=0.84). Extrusion rates also were not found to differ based on whether N-acetylcysteine was infused at the time of insertion[35] or whether the ear was suctioned before tube placement.[41]

The included studies demonstrated a difference in tube retention time that is likely based on tube structure; Goode and Paparella II tympanostomy tubes were retained longer than Armstrong or Shepard tubes. This is consistent with the usual groupings of these types of tubes into shorter-

or longer-acting. Additionally, one study demonstrated that a silicone Shah tube was retained longer than polyethylene tube.[40]

Surgical Interventions: Tympanostomy Tubes Versus Watchful Waiting or Myringotomy

Description of Studies

The evidence consisted of two recent systematic reviews, one of which was a Cochrane Collaboration review by Browning et al. (2010)[17] and the second was by Hellstrom et al., (2011)[15] (Table 9). The Browning et al. review summarized 10 RCTs of tympanostomy tubes in treating children with OME; eight of which were in comparison to watchful waiting or delayed treatment and two were in comparison to myringotomy in the control ear. Hellstrom et al., (2011)[15] reviewed six trials comparing tympanostomy tubes to watchful waiting or myringotomy, all of which were also included in the Cochrane review by Browning et al. We identified two additional RCTs. The first compared tympanostomy tubes to myringotomy in children with conductive hearing loss and tympanostomy tubes to myringotomy or no surgery in children with normal hearing, at three year followup.[44] The second compared tympanostomy tubes to laser myringotomy at 6 month followup.[45] The Browning et al. systematic review was assessed as having a low risk of bias.

Table 9. Characteristics of studies: Tympanostomy tubes versus watchful waiting or myringotomy

Study, Study Type Country	Arm (N)	Diagnosis Criteria	Wait Period Between Diagnosis and Study	Inclusion/ Exclusion Criteria	Length of Study Followup	Age	Risk of Bias
Browning et al., 2010[17] Systematic Review International	Arms differ across comparisons: 10 trials (1728 participants)	Combination of otoscopy (including pneumatic and microscopic), tympanometry and audiometry.	NR	Include: RCTs of short-term tube; randomization could be by child or by ear Exclude: Observational studies or NRCTs; studies including adenoidectomy (unless the arms with adenoidectomy could be excluded)	Child:6-9 mos; 12 mos Ear: 4-6 mos; 7-12 mos	1-12 yrs	Low
Hellstrom et al., 2011[15] Systematic Review International	Arms differ across comparisons: 9 trials (1249 participants)	Swedish guidelines for diagnosis of OME as a painless inflammation with effusion in the middle ear with impaired hearing ≥ 3 mos; no other criteria specified	Minimum of 3 mos	Include: RCTs (individual or ear), NRCTs, and cohort studies published between 1966 and 2007 of efficacy of tubes on hearing, language development, and quality of life and of complications	Various	Children or adolescents	Medium

Table 9. Characteristics of studies: Tympanostomy tubes versus watchful waiting or myringotomy (continued)

Study, Study Type Country	Arm (N)	Diagnosis Criteria	Wait Period Between Diagnosis and Study	Inclusion/ Exclusion Criteria	Length of Study Followup	Age	Risk of Bias
Koopman et al., 2004[45] RCT by ear Amsterdam	G1: TT + cold knife myringotomy (N=208) G2: Laser myringotomy (N=208)	Binocular otoscopy, tympanometry and audiometry	3 mos	Include: Bilateral OME; <11 yrs; 3 mos of hearing problem per parent report Exclude: Unilateral OME; uncooperative; clinically admitted patients, asymmetric perceptive hearing loss; previously operated ears with other than myringotomy or tube	6 mos	< 11 yrs	Medium
Mandel et al., 1989[44] RCT by cluster United States	Without "significant" hearing loss G1:Myringtomy (N=27) G2: Myringotomy + TT (N=30) G3: No surgery (N=39) With significant hearing loss G4: Myringotomy (N=12) G5: Myringotomy + TT (N=11)	Tympanometry and middle-ear muscle reflex testing	2 mos and medical tx	Include: MEE ≥ 2 mos duration persisting after at 1 14-day course of antimicrobial and pseudoephedrine Exclude: craniofacial malformations; systemic illnesses; hx of ear surgery	3 yrs	7 mos to 12 yrs	Medium

Abbreviations: Hx = history; meds = medications; MEE= middle ear effusion; Mos = months; NRCTs = nonrandomized controlled trials; OME = otitis media with effusion; RCT = randomized controlled trial; tube = tympanostomy tubes; tx = treatment; yrs = years.

Key Points

- Tympanostomy tube placement decreased time with persistent middle ear effusion at long term outcomes based on meta-analysis results; 32 percent over 1 year and 13 percent over 2 years. Strength of evidence was high at one year, based on large magnitude of effect, and moderate at two years (Table 10).
- Hearing outcomes were improved with tympanostomy tube placement for up to nine months. Strength of evidence was moderate. In contrast, at longer periods of followup, hearing differences become progressively smaller and were not significantly different at 7

to 12 month (low strength of evidence) and 12 to 18 month (moderate strength of evidence) follow-up periods. All findings were based on meta-analyses.

- We found one small RCT measuring AOM outcomes at 3 years that found no difference between groups. Strength of evidence is insufficient.
- We found no evidence concerning vestibular or health care use outcomes. Strength of evidence is insufficient for no evidence.

Table 10. Strength of evidence for KQ 1: Clinical outcomes and health care utilization

Comparison (G1 vs. G2)	OME Signs and Symptoms	Objective Hearing	AOM	Balance	Health Care Utilization
TT vs. watchful waiting or myringotomy	**High** 32% less time with TT at 1 year **Moderate** 13% less time with TT at 2 years **Insufficient** One study (N=119) time with OME at 3 years, no difference **Insufficient** OME recurrence (No studies)	**Moderate** Better hearing with TT at 4-6 months and 6-9 months **Low** No difference in hearing, by ear, at 7-12 months **Moderate** No difference in hearing, by child, at 12 and 18 months **Insufficient** One study by ear (N=72) at 24 months	**Insufficient** 1 Study	**Insufficient** (No studies)	**Insufficient** (No studies)

Abbreviations: MEE = middle ear effusion; OME = otitis media with effusion; TT = tympanostomy tubes.

Detailed Synthesis

Duration of Middle Ear Effusion

Browning demonstrated a lower likelihood of persistent middle ear effusion in children who received tympanostomy tubes instead of watchful waiting or myringotomy; at 1year post tubes, 32 percent less time (95% CI, 17% to 48%), based on a meta-analysis of three studies and 13 percent at 2 years (95% CI, 8% to 17%), based on a meta-analysis of three studies (Table 11).[17] Similarly, a recent RCT demonstrated less persistent middle ear effusion in subjects who had tympanostomy tubes versus laser myringotomy at monthly followup, at 1 through 6 months ($p<0.001$, at all time points).[45] An earlier RCT by Mandel et al. (1989) also found significantly lower rates of persistent effusion in children who had received tympanostomy tubes compared with myringotomy at 1year($p<0.001$), but results were no longer significant at 2 and 3 year followup.[44]

Table 11. Clinical outcomes: Tympanostomy tubes versus watchful waiting or myringotomy

Study	Arm (N)	Study Duration Until Outcome Measurement	Middle Ear Effusion/ Time With Effusion	OME Recurrence/ Ventilation	AOM	Measured Hearing
Browning et al., 2010[17]	By child (1 study) (N=215)	3 months	NR	NR	NR	Bilateral tube vs. WW: Mean Diff: -11.9 (95% CI, -9.6, -14.2) (favors tube
	(MA:3 studies) (N=523)	6-9 months	NR	NR	NR	Bilateral tube vs. WW: Mean Diff: - 4.20 (95% CI, -6.00 to -2.39) (favors tube)
	(MA: 3 studies) (N=574)	12 months	Bilateral tube vs. myringotomy, delayed treatment or WW: Mean diff: -0.32 (95% CI, -0.48 to -0.17) (favors tube)	NR	NR	NR
	(MA: 2 studies) (N=328)	12 months	NR	NR	NR	Bilateral tubes vs. watchful waiting: Mean Diff - 0.41 (95% CI, -2.37, 1.54)
	MA: 2 studies (N=283)	18 months	NR	NR	NR	Bilateral tube vs. WW Mean Diff: -0.02 (95% CI, -3.22, 3.18)
	MA: 3 studies (N=426)	2 years	Bilateral tube vs. delayed treatment or WW: Mean diff: -0.13 (95% CI, -0.17, -0.08) (favors tube)	NR	NR	NR
	By ear MA: 3 studies (N=230 ears)	4 to 6 months	NR	NR	NR	Unilateral tube vs. WW (2 studies) or myringotomy (1 study): Mean Diff:-10.08 (95% CI, -19.12, -1.05) (favors tube
	MA: 3 studies (N=234 ears)	7 to 12 months	NR	NR	NR	Unilateral tube vs. WW (2 studies) or myringotomy (1 study): Mean Diff: -5.18 (95% CI, -10.43, 0.07) (favors tube)
	1 study (N=72 ears)	24 months	NR	NR	NR	Unilateral tube vs. myringotomy: Mean Diff: -2.1 (95% CI, 2.6, -6.8)

Table 11. Clinical outcomes: Tympanostomy tubes versus watchful waiting or myringotomy (continued)

Study	Arm (N)	Study Duration Until Outcome Measurement	Middle Ear Effusion/ Time With Effusion	OME Recurrence/ Ventilation	AOM	Measured Hearing
Koopman et al., 2004[45]	G1: TT + cold knife myringotomy (N=208) G2: Laser myringotomy (N=208)	1 month	Absence of effusion G1:87.4% G2:46.6%	NR	NR	NR
		2 months	G1:81.9% G2:35.5%	NR	NR	NR
		3 months	G1:81.5% G2:38.6%	NR	NR	NR
		4 months	G1:75.5% G2:41.6%	NR	NR	NR
		5 months	G1:68.5% G2:39.1%	NR	NR	NR
		6 months	G1:70.7% G2:39.1% (all p<0.001[a])	NR	NR	NR
Mandel et al., 1989[44]	Without "significant" hearing loss: G1:Myringtomy G2: Myringotomy + TT G3: No surgery With "significant hearing" loss: G4: Myringotomy G5: Myringotomy + TT	1 year	% Time with OME G1: 56.6% G2:16.4% G3:56.3% G4:56.7% G5:9.8% (G1 or G3 vs. G2: p<0.001) (G4 vs. G5: p<0.001)	NR	NR	SRT in dB (2 months. post tx) Right ear G1: 18.5 G2: 16.2 G3: 6.2 G4: 22.0 G5: 5.5
	NR	2 years	G1:35.2% G2:20.4% G3:28.2% G4:39.9% G5:28.3% (p=NS)	NR	NR	NR
	NR	3 years	G1: 25.5% G2: 25.0% G3:19.2% G4: 14.4% G5: 30.3% (p=NS)	NR	Episodes/ person-year) G1: 0.58 G2: 0. 18 G3: 0.38 G4: 0.31 G5: 0.41	NR

[a]p values calculated by investigators

Abbreviations: AOM = acute otitis media; MA = meta-analysis; NR = not reported; NS = not significant; OME = otitis media with effusion; SRT = speech related threshold; TT = tympanostomy tube; WW = watchful waiting; vs.= versus.

Measured Hearing

Meta-analysis results of RCTs presented in the Browning et al. review (2010) showed a significant improvement in hearing in tympanostomy tube arms compared with watchful waiting or myringotomy in the short term but not after longer periods of followup.[17] More specifically, meta-analysis results of three studies randomized by ear, at 4 to 6 month followup, found mean differences in hearing to be -10.18 dB (95% CI, -19.12 to -1.05). (Negative results represent

hearing improvement in the tympanostomy tube group.) Meta-analysis results from three studies randomized by child also demonstrated significantly better hearing in the tympanostomy group at 6 to 9 month followup, -4.20 dB (95% CI, -6.00 to -2.39). However, hearing levels were no longer significantly different at 7 to 12 month followup based on meta-analysis results of three "by ear" studies, -5.18 dB (-10.43 to 0.07) and two meta-analyses of RCTs randomized by child, at 12 months and 18 months, both based on two studies (-0.41dB [95% CI, -2.37 to 1.54] and -0.02 dB [95% CI, -3.22 to 3.18], respectively).[17]

Recurrent AOM

After 3 years of observation, even though rates in all groups were low, children who had received tympanostomy tubes had the lowest relative rates of AOM after placement (significance not presented)[44].

Surgical Interventions: Tympanostomy Tubes Plus Adenoidectomy Versus Myringotomy Plus Adenoidectomy or Adenoidectomy Alone

Description of Studies

No systematic reviews were found that evaluated outcomes from tympanostomy tubes in addition to adenoidectomy. We identified eight studies that included adenoidectomy in both intervention groups and tympanostomy tubes in at least one of the intervention groups (Table 12).[46-53] Three of the studies compared tympanostomy tubes to controls by ear, among children who all had adenoidectomies.[46-48] The other five studies compared tympanostomy tubes to myringotomy, among children who all had adenoidectomies.[49-53] Length of study followup generally ranged from 2 days to 12 months. However, one study followed patients for 7 years.[53]

Table 12. Characteristics of studies: Tympanostomy tubes and adenoidectomy versus watchful waiting or myringotomy and adenoidectomy or adenoidectomy alone

Study, Study Type, Country	Arm (N)	Diagnosis Criteria	Wait Period Between Diagnosis and Study	Inclusion/ Exclusion Criteria	Length of Study Followup	Age	Risk of Bias
Brown, Richards, and Ambegaokar, 1978[46] RCT by ear Wales	G1: TT + adenoidectomy (N=55) G2: Adenoidectomy (N=55)	History, Otoscopy, Audiometry	Not specified	Not specified	48 hrs, 3 6, 9, 12 mos, 5 yrs	4-10 yrs	Medium

Table 12. Characteristics of studies: Tympanostomy tubes and adenoidectomy versus watchful waiting or myringotomy and adenoidectomy or adenoidectomy alone (continued)

Study, Study Type, Country	Arm (N)	Diagnosis Criteria	Wait Period Between Diagnosis and Study	Inclusion/ Exclusion Criteria	Length of Study Follow-up	Age	Risk of Bias
Austin, 1994[47] NRCT by ear United States	G1: TT+ adenoidectomy (N=31) G2: Adenoidectomy (N=31)	Audiometry	Not specified	Include: Indication for adenotonsillectomy and OME, Resistant to ENT or pediatric management	3 mos	Not reported	Medium
Lindholdt, 1977[48] NRCT by ear Denmark	G1: TT+ adenoidectomy (N=91) G2: Adenoidectomy (N=91 ears	Tympanometry	Not specified	Include: Bilateral OME; minimal differences between ears in pressure and hearing Exclude: Previous ear surgery	Until extrusion ,18 mos	Mean: 4 yrs (range 1-10 yrs)	Medium
D'Eredita and Shah, 2006[49] RCT by person Italy	G1: CDLM + adenoidectomy (N=15) G2: TT + adenoidectomy (N=15)	Tympanometry	3 mos	Include: 3 mos of OME Exclude: Hx of prior surgery, craniofacial syndrome, MR or cognitive disorder	12 mos	4 yrs (range 2-6 yrs)	Medium
Popova et al., 2010[50] RCT by person Bulgaria	G1: TT + adenoidectomy (N=42) G2: Myringotomy + adenoidectomy (N=36)	Pneumatic otoscopy and tympanometry	3 mos	Include: OME for 3 mos; conductive hearing loss >20dB Excluded: Previous ear or throat surgery; craniofacial syndromes; destructive middle ear disease; conductive hearing loss attributed to destructive middle ear changes; sensorineural hearing loss	12 mos	G1: 60 mos G2: 61 mos	Medium
Shishegar and Hobhoghi, 2007[51] RCT by ear Iran	G1: TT + adenoidectomy (n=30 ears) G2: Myringotomy + adenoidectomy (N=30 ears)	Otoscopy, tympanometry, audiometry	NR	Include: Bilateral OME unresponsive to medical therapy Exclude: Prior ear surgery; prior adenoidectomy; no OME, cleft palate; perforated TM	6 mos	Range: 4-8 yrs	Medium

Table 12. Characteristics of studies: Tympanostomy tubes and adenoidectomy versus watchful waiting or myringotomy and adenoidectomy or adenoidectomy alone (continued)

Study, Study Type, Country	Arm (N)	Diagnosis Criteria	Wait Period Between Diagnosis and Study	Inclusion/ Exclusion Criteria	Length of Study Followup	Age	Risk of Bias
Vlastos et al., 2011[52] RCT by person Greece	G1: Adenoidectomy + TT (N=25) G2: Adenoidectomy + Myringotomy (N=27)	Otoscopy, tympanometry, and pure tone audiometry	NR	Include: Scheduled for adenoidectomy due to sleep apnea; bilateral OME; >3 yrs age Excluded: No OME; chronic OME; previous ear surgery; language delays; behavioral problems; anatomic changes	12 mos	G1: 4.6 yrs (range 3-7 yr) G2: 4.4 yrs (range 3-7 yr)	Medium
Tos and Stangerup, 1989[53] NRCT by ear Denmark	G1: TT + adenoidectomy (N=146) G2: Myringotomy + adenoidectomy (N=146)	NR	NR	Include: Bilateral OME	7 yrs	5 yrs	Medium

Abbreviations: dB = decibel; hrs = hours; Hx = history; mos = months; MR = mental retardation; NR = not reported; NRCT = nonrandomized controlled trial; OME = otitis media with effusion; RCT = randomized controlled trial; TM = tympanic membrane; TT = tympanostomy tubes; yrs = years; CDLM = Contact diode laser myringotomy

Key Points

- We found no differences in short-term and long-term hearing outcomes, when tympanostomy tubes or myringotomy were added as treatment in addition to adenoidectomy (Table 13). Strength of evidence was low.
- Evidence was insufficient in relation to differences in reoccurrence of OME.
- Evidence was insufficient for AOM, vestibular outcomes and health care service use because we found no studies.

Table 13. Strength of evidence for KQ 1: Clinical outcomes and health care utilization

Comparison (G1 vs. G2)	Middle ear effusion/ Time with effusion	OME Recurrence/ Ventilation	AOM	Measured Hearing
TT+adenoidectomy vs. Adenoidectomy alone	**Insufficient** Single study No difference	**Insufficient** (No studies)	**Insufficient** (No studies)	**Insufficient** No difference, single small study
TT+adenoidectomy vs. Myringotomy + adenoidectomy	**Insufficient** (No studies)	**Insufficient** Mixed results	**Insufficient** (No studies)	**Low** No difference (6 mos, 12 mos and >3 years)

Abbreviations: mos = months; TT = tympanostomy tubes; yrs = years

Detailed Synthesis

Recurrence of Middle Ear Effusion

Two studies found conflicting results regarding middle ear ventilation or recurrence of OME. Results from one study indicated that the middle ear was likely to be ventilated longer after

tympanostomy tube placement than myringotomy when all individuals received adenoidectomy (6.3 months versus 3.5 months respectively, p<0.001) (Table 14).[49] A second study in which all participants received an adenoidectomy failed to find a difference in recurrence of OME between ears that received tubes and ears that received myringotomy.[50] Also, middle ear fluid reoccurrence was not significantly different among children with adenoidectomies who also received tympanostomy tubes in comparison to no additional surgery, at 5 year followup.[46]

Table 14. Clinical outcomes: Tympanostomy tubes plus adenoidectomy versus watchful waiting or myringotomy plus adenoidectomy or adenoidectomy alone

Study	Arm (N)	Study Duration Until Outcome Measurement	Middle Ear Effusion/ Time With Effusion	OME Recurrence/ Ventilation	AOM	Measured Hearing
Brown et al., 1978[46]	G1: TT + adenoidectomy (N=55) G2: Adenoidectomy (N=55)	48 hrs 3 mos 5 yrs	NR	G1: 2% G2: 4% No "significant difference" in recurrence between groups	NR	PTA G1: 8.9 dB* G2: 24.7 dB G1: 11.4dB* G2: 16.6 dB G1: 17 dB G2: 14 dB *(significant but no p-value reported)
Austin, 1994[47]	G1: TT+ adenoidectomy (N=31 ears) G2: Adenoidectomy (N=31 ears)	1-3 mos	NR	NR	NR	Air-bone gap G1: 13.2 G2: 14.4 p>0.1 Mean improvement in Air-bone gap G1:16 dB G2: 12.2 dB p>0.1 Mean Difference Between tx: 1.9 dB
D'Eredita and Shah, 2006[49]	G1: CDLM + adenoidectomy (N=15)	NR	NR	Middle ear ventilation maintained G1: 3.5 mos G2: 6.3 mos p<0.001	NR	"Normal in both groups at 1 year followup"
	G2: TT + adenoidectomy (N=15)	3 mos	NR	Number ears ventilated (%) G1: 11 (36.6) G2: 30 (100) p=NR	NR	NR
Popova et al., 2010[50]	G1: TT + adenoidectomy (N=42) G2: Myringotomy + adenoidectomy (N=36)	1 mos 6 mos	NR	NR	NR	PTA G1: 13.9 dB G2: 14.1 dB p=0.83 G1: 7.6 dB G2: 8.0 dB p=0.68
	NR	12 mos	NR	OME recurrence G1: 14% G2: 10% p=0.547		G1: 5.5 dB G2: 6.3 p=0.24

Table 14. Clinical outcomes: Tympanostomy tubes plus adenoidectomy versus watchful waiting or myringotomy plus adenoidectomy or adenoidectomy alone (continued)

Study	Arm (N)	Study Duration Until Outcome Measurement	Middle Ear Effusion/ Time With Effusion	OME Recurrence/ Ventilation	AOM	Measured Hearing
Shishegar and Hobhoghi, 2007[51]	G1: TT + adenoidectomy (n=30 ears) G2: Myringotomy + adenoidectomy (N=30 ears)	1 mos	NR	NR	NR	Air-bone gap improvement G1: 17.47 dB G2: 16.04 dB (p=ns) Mean SRT Hearing threshold G1: 18.3 G2: 17 dB (p=ns)
	NR	6 mos	NR	NR	NR	Air-bone gap improvement G1:17.62 dB G2: 16.25 dB (p=ns) Means SRT Hearing threshold G1: 19.3 dB G2: 17.16 dB (p=ns)
Vlastos et al., 2011[52]	G1: TT + Adenoidectomy (N=25) G2: Myringotomy + Adenoidectomy (N=27)	6 mos	NR	NR	NR	Change in Hearing G1: -7.41 G2: -4.06 Mean HL Change 3.35 dB (95% CI, -6.64 to 10.35)
	NR	12 mos	NR	NR	NR	Change in Hearing G1: -8.06 dB G2: -7.40 dB Mean HL Change 0.66 dB(95% CI, -6.82 to 8.15)

Table 14. Clinical outcomes: Tympanostomy tubes plus adenoidectomy versus watchful waiting or myringotomy plus adenoidectomy or adenoidectomy alone (continued)

Study	Arm (N)	Study Duration Until Outcome Measurement	Middle Ear Effusion/ Time With Effusion	OME Recurrence/ Ventilation	AOM	Measured Hearing
Tos and Stangerup, 1989[53]	G1: TT + adenoidectomy (N=146) G2: Myringotomy + adenoidectomy (N=146)	2-3 yrs	NR	NR	NR	PTA G1: 15.0 dB G2: 14.7 dB (p not reported) Gain after tx G1: 14.5 dB G2: 13.1 dB (p not reported)
		6-7 yrs	NR	NR	NR	PTA G1: 11.7 dB G2: 11.1 dB (p not reported) Gain after tx G1: 3.3 dB G2: 3.6 dB

Abbreviations: AOM = acute otitis media; dB = decibel; fU = followup; HL = hearing level; hrs = hours; mos = month; NR = not reported; OME = otitis media with effusion; PTA = pure tone audiometry; RCT = Randomized controlled trial; SRT = speech reception threshold; TT = tympanostomy tubes; tx = treatment; yrs = years.

Measured Hearing

Hearing outcomes were compared in two studies of tympanostomy tubes plus adenoidectomy versus adenoidectomy alone.[46, 47] Followup ranged from 2 days to 5 years. One of the studies found no difference in hearing at the post-operative assessment.[47] A second study[46] found hearing improvement during the first 3 months in the tympanostomy tube group, but by 5 years hearing levels were similar.[46]

In five studies that evaluated tympanostomy tubes versus myringotomy in children who had adenoidectomy, hearing was not significantly different at any time point, ranging from 1month to 6 year followup.[49-53]

Surgical Interventions: Myringotomy Comparisons

Description of Studies

The included evidence consisted of one RCT by Ragab[54] (Table 15). This study was designed to compare two different procedures for myringotomy; namely radio frequency myringotomy with Mitomycin C, a topical chemotherapeutic agent, in comparison to radio frequency myringotomy alone. In this trial, a subset of individuals received an adenoidectomy (73% and 67% respectively by arm). Followup was short-term.

Table 15. Characteristics of studies: Myringotomy comparisons

Study, Study Type, Country	Arm (N)	Diagnosis Criteria	Wait Period Between Diagnosis and Study	Inclusion/ Exclusion Criteria	Length of Study Follow-up	Age	Risk of Bias
Ragab, 2005[54] RCT Egypt	G1: Radiofrequency myringotomy + Mitomycin C (N=30) G2: Radiofrequency myringotomy (no Mitomycin C) (N=30)	History, pneumatic otoscopy and tympanometry	NR	Include: Patients undergoing surgery for OME	3 mos	G1: 4.8 yr G2: 5.2 yr	Medium

Abbreviations: mos = months = OME = otitis media with effusion; RCT = randomized control trial; yr = years;

Key Points

- One small RCT comparing approaches to myringotomy found a significant difference in resolution of OME favoring myringotomy with Mitomycin C but no significant differences in hearing improvement. Based on one small study, the evidence is graded as insufficient.

Detailed Synthesis

OME Signs and Symptoms Outcomes

Ragab[54] examined resolution of middle ear effusion and reported a significant difference favoring radio frequency myringotomy with Mitomycin C ($p<0.01$) (Table 16). This study did not present data on either OME recurrence or AOM.

Table 16. Clinical outcomes: Myringotomy comparisons

Study	Arm (N)	Study Duration Until Outcome Measurement	Middle Ear Effusion/ Time With Effusion	Measured Hearing
Ragab, 2005[54]	G1: Radiofrequency myringotomy + Mitomycin C (N=30) G2: Radiofrequency myringotomy (no Mitomycin C) (N=30)	3 months	Resolution of OME G1: 59% G2: 28% p<0.01	Air Bone Gap Improvement: G1: 12 dB G2: 10 dB G1 vs. G2 p=NS

Abbreviations: AOM = acute otitis media; dB = decibel; OME = otitis media with effusion.

Hearing Outcomes

Both myringotomy with and without Mitomycin C groups demonstrated a significant air-bone gap improvement 3 months post surgery as compared with pre-surgery but there was no significant difference in improvement in air-bone gap improvement rates between the two groups.[54]

Surgical Interventions: Myringotomy with Adenoidectomy Comparisons

Description of Studies

One retrospective cohort study from the United States compared two different procedures for myringotomy—namely, laser myringotomy with cold knife myringotomy (Table 17).[55] In both arms, all participants received an adenoidectomy. Patients included children older than 4 years of age who had refractory OME or children of any age who had a need for a second tube insertion.

Table 17. Characteristics of studies: Myringotomy with adenoidectomy

Study, Study Type, Country	Arm (N)	Inclusion/ Exclusion Criteria	Length of followup	Age in Years (Range)	Risk of Bias
Szeremeta et al., 2000[55] Retrospective cohort USA	G1: Laser myringotomy with adenoidectomy (N=39) G2: Cold knife myringotomy with adenoidectomy (N=48)	Include: Children >4 yrs with refractory OME or with a need for a 2nd tube; spring operations	Mean time in months (range) G1: 16.6 (6–27) G2: 20.2 (12–48)	Mean: G1: 6.5 (2.74 to 12.52) G2: 7.4 (3.86 to 5.34)	Medium

Abbreviations: OME = otitis media with effusion; yrs = years.

Key Points

- A particular approach to myringotomy (laser versus cold knife) among patients who had also all received adenoidectomy displayed mixed findings in relation to clinical outcomes. The study did not find that laser myringotomy was superior to cold knife myringotomy in the percentage of patients with OME but did find a difference in the patency of ears (open hole based on myringotomy), post-operatively. Based on one small study, the evidence is graded as insufficient.
- This study did not report any other clinical or health care utilization outcomes such as AOM, balance, or use of health care services.

Detailed Synthesis

Outcomes focused on the percentage of ears among children with middle ear effusion and patency of ears at the first post-surgery visit (Table 18). Laser myringotomy with adenoidectomy did not differ from cold knife myringotomy with adenoidectomy in the percentage of ears presenting with middle ear effusion at followup. However, the authors reported a significant difference in the percentage of ears that were patent at the first post-operative visit, favoring laser myringotomy with adenoidectomy ($p<0.01$).

Table 18. Clinical outcomes: Myringotomy with adenoidectomy comparisons

Study	Arm (N)	Study Duration Until Outcome Measurement	Middle Ear Effusion/Time With Effusion	OME Recurrence Or Ventilation
Szeremeta et al., 2000[55]	G1: Laser myringotomy with adenoidectomy (N=39) G2: Cold knife myringotomy with adenoidectomy (N=48)	Within 50 days	MEE G1: 10% G2: 15% p>0.1	Patency G1: 21% G2: 0% p<0.01

Abbreviations: MEE = middle ear effusion; OME = otitis media with effusion.

Surgical Interventions: Adenoidectomy Versus Other Interventions

Description of Studies

All of the evidence comparing the effectiveness of adenoidectomy to non-surgical interventions was obtained from a recent Cochrane Collaboration systematic review by van den Aadweg et al.[18] (Table 19). Included studies were limited to RCTs of non-infant children (2-14 years of age) with persistent OME who were followed for 6 months or more. The review included seven studies limited to OME patients (N=1,177). Treatment comparisons included: adenoidectomy with and without myringotomy versus non-surgical treatment or myringotomy only; adenoidectomy with unilateral tympanostomy tubes versus a unilateral tympanostomy tube only (comparison by ears); and adenoidectomy with bilateral tympanostomy tubes versus bilateral tympanostomy tubes only.

Table 19. Characteristics of studies: Adenoidectomy versus other interventions

Study, Study Type, Country	Arm (N)	Diagnosis Criteria	Wait Period Between Diagnosis and Study	Inclusion/ Exclusion Criteria	Length of Study Followup	Age Range	Risk of Bias
van den Aardweg et al., 2010[18] Systematic review International	Arms differ across comparisons: 7 trials (1,177 participants)	Various criteria including clinical judgment, otoscopy, tympanometry, pure tone thresholds	Various	Include: RCTs of adenoidectomy for otitis media compared with non-surgical tx or TT alone; children ≤18 yrs of age Exclude: Quasi randomized trials (e.g., allocation by DOB or record number)	At least 6 mos	2-14 yrs of age	Low

Abbreviations: DOB = date of birth; mos = months; RCT = Randomized controlled trial; tx = treatment; yrs = years.

Key Points

- Adenoidectomy was superior to both unilateral tympanostomy tube placement and watchful waiting to resolve OME at both 6 months and 12 months followup (Table 20).[18] At 6 months, the risk difference was 0.27 (95% CI, 0.13 to 0.42), measured through otoscopy and 0.22 (95% CI, 0.12 to 0.32) measured through tympanometry. At 12 months the risk difference was 0.29 (95% CI, 0.19 to 0.39). High strength of evidence.

- Resolution of OME and hearing were superior with adenoidectomy and myringotomy at 24 months in one large study (N=237) compared to myringotomy alone. Strength of evidence was low.
- OME resolution outcomes were mixed in studies examining whether the addition of tympanostomy tubes was superior to adenoidectomy alone. Strength of evidence was insufficient
- Evidence was insufficient to determine the comparative effectiveness of adenoidectomy versus no treatment or tympanostomy tubes in relation to hearing, vestibular, and health services related outcomes.

Table 20. Strength of evidence for KQ 1: Clinical outcomes and health care utilization

Comparison (G1 vs. G2)	OME Signs and Symptoms	Objective Hearing	AOM	Balance	Health Care Utilization
Adenoidectomy vs. no treatment or tympanostomy tubes	**High** OME resolution favors adenoidectomy vs. no treatment 2 MA at 6 mo, 1 MA at 12 mos	**Insufficient** Mixed results	**Insufficient** Mixed results	**Insufficient** No studies	**Insufficient** No studies
Adenoidectomy vs. myringotomy	**Low** Mean time with effusion favors adenoidectomy+ myringotomy over myringotomy alone at 24 mos, 1 RCT (N=237)	**Low** Hearing favors adenoidectomy+ myringotomy over myringotomy alone at 24 mos, 1 RCT (N=237)	**Insufficient** No studies	**Insufficient** No studies	**Insufficient** No studies

Abbreviations: OME = otitis media with effusion; AOM = acute otitis media; mos = months; MA = meta-analysis

Detailed Synthesis

OME Signs and Symptoms Outcomes

OME resolution was superior in the adenoidectomy group compared with no treatment in three meta-analyses that compared unoperated ears in patients with unilateral tympanostomy tubes, both at 6 months based on otoscopy (risk difference=0.27, based on 2 studies) and tympanometry (risk difference=0.22, based on 3 studies) and at 12 months based on tympanometry (risk difference=0.29, based on 3 studies) (Table 21).[18] We found that results were mixed across studies concerning whether the addition of tympanostomy tubes to adenoidectomy improved OME-related outcomes.

Table 21. Clinical outcomes: Adenoidectomy versus other interventions

Study	Arm (N)	Study Duration Until Outcome Measurement	MEE/Time With Effusion	OME Resolution	AOM	Measured Hearing
van den Aardweg et al., 2010[18]	1 study Adenoidectomy + unilateral TT vs. unilateral TT (N=72)	6 mos	NR	NR	NR	Mean hearing level (dB) Ad; 18.0 No tx: 21.1 SMD: -0.25 (95% CI, -0.71 to 0.22)
		12 mos	NR	Otoscopy Ad: 54% No tx: 37% Risk diff: 17% (95% CI, -6% to 40%)		Ad; 15.6 No tx: 18.4 SMD: -0.29 (95% CI, -0.76 to 0.17)
	1 study Adenoidectomy + unilateral TT vs. unilateral TT (N=81)	6 mos	NR	NR	NR	Mean hearing level (db) Ad: 20.4 No tx: 36.5 SMD: -1.37 (95% CI, -1.87 to -0.88)
		12 mos	NR	Otoscopy Ad: 69.4% No tx: 27.7% Risk diff: 42% (95% CI, 22% to 62%)	NR	Ad; 19.7 No tx: 27.4 SMD: -0.67 (95% CI, -1.12 to -0.22)
	1 study Adenoidectomy + Myringotomy vs. Myringotomy (N=88)	6 mos	NR	Normal ears (Type A tympanogram) Ad + M: 68% M: 52% Risk diff: 15% (95% CI, -5% to 46%)	NR	NR
	1 study Adenoidectomy + Myringotomy vs. Myringotomy (N=237)	24 mos	Mean time with effusion Ad + M: 0.302 M: 0.491 SMD: -0.76 (95% CI, -1.02 to -0.49)	NR	NR	Mean time HL $\geq$ 20 dB better ear Ad + M: 0.078 M: 0.186 SMD: -0.66 (95% CI, -0.93 to -0.40) Mean time HL $\geq$ 20 dB worse ear Ad + M: 0.220 M: 0.375 SMD: -0.65 (95% CI, -0.91 to -0.39)
	MA: 2 studies Adenoidectomy + unilateral TT vs. unilateral TT (N =153)	6 mos	NR	Otoscopy Ad + uni TT: 49% uni TT: 21% Risk diff: 0.27 (95% CI, 0.13 to 0.42)	NR	NR

Table 21. Clinical outcomes: Adenoidectomy versus other interventions (continued)

Study	Arm (N)	Study Duration Until Outcome Measurement	MEE/Time With Effusion	OME Resolution	AOM	Measured Hearing
van den Aardweg et al., 2010[18] (continued)	MA; 3 studies Adenoidectomy + unilateral TT vs. unilateral TT (N =297)	6 mos	NR	Tympanometry Ad + uni TT: 39% Uni TT: 17% Risk diff: 0.22 (95% CI, 0.12 to 0.32)	NR	NR
	MA; 3 studies Adenoidectomy + unilateral TT vs. unilateral TT (N =298)	12 mos	NR	Tympanometry Ad + uni TT: 47% uni TT: 20% Risk diff: 0.29 (95% CI, 0.19 to 0.39)	NR	NR
	1 study Adenoidectomy + bilateral TT vs. bilateral TT (N = 95)	12 mos	% with effusion Ad + bil TT: 18% bil TT: 23% Risk diff: -5% (95% CI, -8% to 17%)	NR	NR	NR
		24 mos	Ad + bil TT: 15% bil TT: 18% Risk diff: -3% (95% CI, -10% to 15%)	NR	NR	NR
	1 study Adenoidectomy + bilateral TT vs. bilateral TT (N=254)	24 mos	Mean time with effusion Ad + bil TT: 0.258 bil TT: 0.349 SMD: -0.40 (95% CI, -0.65 to -0.15	NR	NR	Mean time HL ≥ 20 dB better ear Ad + bil TT: 0.065 Bil TT: 0.101 SMD: -0.23 (95% CI, -0.48 to 0.02) Mean time HL ≥ 20 dB worse ear Ad + bil TT: 0.224 Bil TT: 0.304 SMD: -0.35 (95% CI, -0.60 to -0.11)
	1 study Adenoidectomy + bilateral myringotomy + unilateral TT vs. bilateral myringotomy + unilateral TT (N=149)	6 mos	NR	NR	NR	Change in mean dB Diff between Ad + bil M + uni TT vs.. bil M + uni TT: 4.3 (95% CI, 1.4 to 9.9)
	1 study Adenoidectomy + bilateral myringotomy + unilateral TT vs. bilateral myringotomy + unilateral TT (N=149)	12 mos	NR	NR	NR	Change in mean dB Diff between Ad + bil M + uni TT vs. bil M+ uni TT: 4.3 (95% CI, - 3.1 to 11.6)

Table 21. Clinical outcomes: Adenoidectomy versus other interventions (continued)

Study	Arm (N)	Study Duration Until Outcome Measurement	MEE/Time With Effusion	OME Resolution	AOM	Measured Hearing
van den Aardweg et al., 2010[18] (continued)	1 study Adenoidectomy + Myringotomy + bil TT vs. Myringotomy + bil TT (N=62)	18 mos	Mean time with effusion Ad + M + bil TT: 18% M + bil TT:12% Diff: 6% (95% CI, -12 to 24)	NR	NR	NR
	1 study Adenoidectomy + Myringotomy + bil TT vs. Myringotomy + bil TT (N=44)	18 mos	NR	NR	# episodes Ad + M + bil TT: 7 M + bil TT: 6 Risk diff: 5% (95% CI, -22 to 32)	NR
	1 study Adenoidectomy + Myringotomy + bil TT vs. Myringotomy + bil TT (N=62)	36 mos	Mean time with effusion Ad + M + bil TT: 21% M + bil TT:19% Diff: 2% (95% CI, -19 to 23)	NR	NR	NR
	1 study Adenoidectomy + Myringotomy + bil TT vs. Myringotomy + bil TT (N=39)	36 mos	NR	NR	# episodes Ad + M + bil TT: 17 M + bil TT: 21 Risk diff: -18% (95% CI, -37 to 1)	NR

Abbreviations: Ad = adenoidectomy; AOM = acute otitis media; bil = bilateral; dB = decibel; diff = difference; HL = hearing level; M = myringotomy; MA = meta-analysis; mos = months; MEE = middle ear effusion; OME = otitis media with effusion; SMD = standard mean difference; TT = tympanostomy tubes; tx = treatment; uni = unilateral.

Hearing Outcomes

Hearing measures and outcomes differed across studies in comparison to tympanostomy tubes and results were mixed. Adenoidectomy and myringotomy was found to be superior to myringotomy alone in one relatively large RCT (N=237) at 24 months in relation to hearing in both the better and worse ear.[18]

Other Outcomes

Episodes of AOM were measured in one study included in the systematic review and they did not differ at 18 or 36 months. No studies measured vestibular function.

Pharmaceutical Interventions: Oral or Topical Nasal Steroids

Description of Studies

The included evidence consisted of one recent Cochrane Collaboration systematic review[20] and one more recent trial, conducted by Williamson et al.[56, 57] (Table 22). The systematic review summarized evidence from nine RCTs of oral steroids and two RCTs of topical intranasal steroids, excluding studies limited to ears (rather than children). The Williamson et al. study with

topical intranasal steroids and was conducted by the UK Health Technology Assessment Programme, published as a report[57] and peer-reviewed manuscript.[56] All studies were in comparison with placebo controls. Neither included participants older than 12 years of age. The systematic review did not exclude children with comorbidities, whereas the Williamson et al. study excluded children with Down syndrome, cleft palate, and other comorbidities (Table 22). The Cochrane review included one to two and six month followup; the Williamson RCT included 1, 3, and 9 month followup. Both studies were assessed as low risk of bias.

Table 22. Characteristics of studies: Oral or topical nasal steroids

Study, Study Type, Country	Arm (N)	Diagnosis Criteria	Wait Period Between Diagnosis and Study	Inclusion/ Exclusion Criteria	Length of Followup	Age Range	Risk of Bias
Thomas et al., 2010[20] Systematic review International	Arms differs across comparisons: 11 trials (862 participants)	OME: A. Air-bone gap of 10 dB or more + 2 or more of: otomicroscopy, pneumatic otoscopy, tympanometry (type B or C2) B. 2 or more of: otomicroscopy, pneumatic otoscopy, tympanometry (type B or C2) C. 1 of otoscopy alone or tympanometry (type B or C2) D. Poorly or not defined Sig hearing loss defined by: A. Pure-tone audiometry hearing loss of >20 dB at 2 or more times within 3 mos (for example, mean of 500, 1000, and 2000 Hz hearing loss bilaterally) B. Defined, but less strict than A C. Uncertain or not defined	NR	Include: RCTs of oral and topical intranasal steroids, including studies using non-intervention controls with adequate blinding of outcome assessor. Exclude: Observational studies, studies reporting outcomes only with ears as unit of analysis; studies (or data from arms of studies) comparing steroid + additional treatment vs. treatment with placebo + placebo because effect of steroid could not be isolated.	1-2, and 6 months	0-12 years	Low

Table 22. Characteristics of studies: Oral or topical nasal steroids (continued)

Study, Study Type, Country	Arm (N)	Diagnosis Criteria	Wait Period Between Diagnosis and Study	Inclusion/ Exclusion Criteria	Length of Followup	Age Range	Risk of Bias
Williamson et al. 2009[56] Williamson et al. 2009[57] RCT UK	G1: Mometasone furoate nasal spray (N= 105) G2: Placebo spray (N=112)	Tympanometry	In year 1, after failing a second otoscopy screening after 3 mos of watchful waiting. Following protocol change at end of year 1, after failing a second otoscopic screening	Include: Hx of OME by medical record of ≥ 1 episodes in past year or a hx suggestive of OME. At time of study, dx of bil OME. Exclusion: Tympanometry screen passed (A or C1); large amounts of wax; uninterruptable tympanogram; co-morbidities including cleft palate; Down syndrome; primary ciliary dyskinesia; tubes or tympanic perforation, frequent or heavy epistaxis; hypersensitivity to mometasone; history of steroid use in previous 3 mos	1, 3, & 9 mos	4-11 yrs	Low

Abbreviations: db = decibel; bil = bilateral; dx = diagnosis; hx; history; Hz = Hertz; mos = months; NR = not reported; OME = otitis media with effusion; RCT = randomized control trial; yrs = years.

Key Points

- Meta-analyses comparing oral steroids and controls (N=106) did not show differences in middle ear effusion at 1 or 2 months followup (low strength of evidence) (Table 23).
- Meta-analyses comparing oral steroids and controls (N=243) (with oral antibiotic adjunctive therapy) did not show differences in middle ear effusion at 1 or 2 months followup (medium strength of evidence)
- We found insufficient evidence comparing oral steroids with controls (with or without oral antibiotic adjunctive therapy) at 3 months or longer for any hearing outcomes.
- Patients receiving topical intranasal steroids did not differ from controls in cure rates or hearing loss at 3 month or longer followup, based on results from one low risk of bias study (low strength of evidence).
- No studies reported on AOM or other clinical or health care use outcomes (insufficient evidence).

Table 23. Strength of evidence: Clinical outcomes and health care utilization

Comparison	OME Signs And Symptoms	Measured Hearing
Oral steroids vs. control (1-2 mo)	**Low** Persisting OME: no difference MA, 3, 106	**Insufficient** Hearing gain: no diff 1, 49
Oral steroids + antibiotic vs. control + antibiotic (1-2 mo)	**Medium** Persisting OME: no difference MA, 2, 243	**Insufficient** No studies
Topical intranasal steroid vs. control (1, 3 and 9 or more months)	**Low** Cure rate: no diff 1, 217	**Low** Hearing loss: no diff 1, 217
Topical intranasal steroid + antibiotic vs. control + antibiotic (3 or more months)	**Insufficient** Persisting OME (6 mo): no diff 1, 59	**Insufficient** No studies
Oral steroids vs. control (3 months)	**Insufficient** No studies	**Insufficient** Hearing gain: no difference 1, 49
Oral steroids + antibiotic vs. control + antibiotic (6 or more months)	**Insufficient** Persisting: No diff 1, 15	**Insufficient** No study

Detailed Synthesis

OME Signs and Symptoms Outcomes

The Cochrane systematic review[20] and one newer study from our search[56, 57] presented results on outcomes related to middle ear effusion through two measures: persisting OME and cure rates as measured by a flat tympanogram (Table 24). We found no differences in any treatment comparisons at any end points. At 1to 2 month followup, the systematic review found no difference between oral steroids versus controls in relation to persisting OME, OR=0.55 (95% CI, 0.21 to 1.48) based on a meta-analysis of three studies (N=106) or similarly for oral steroids versus controls, when both arms also received antibiotic treatment, OR=0.75 (95% CI, 0.45 to 1.27), based on a meta-analysis of three studies (N=243).[20]The Williamson et al. study also found no significant difference in cure rates in topical steroids versus controls at 1month, controlling for season, age, atrophy, and clinical severity, RR=0.97 (95% CI, 0.74 to 1.26) (N=194).

Table 24. Clinical outcomes: Oral or topical nasal steroids

Study	Arm (N)	Study Duration Until Outcome Measurement	Middle Ear Effusion/ Time With Effusion	Measured Hearing
Thomas et al., 2010[20]	MA: 3 studies (N=106)	1-2 months	Persisting OME Oral steroids vs. control Peto OR: 0.55 (95% CI, 0.21 to 1.48)	NR
	1 study (N=49)	NR	NR	Hearing gain by at least 10 dB Oral steroids vs. control OR: 1.47 (95% CI, 0.39 to 5.57) (baseline: NR)
	MA: 2 studies (N=243)	NR	Oral steroids + antibiotic vs. control + antibiotic Peto OR: 0.75 (95% CI, 0.45 to 1.27)	NR
	1 study (N=59)	3 mos	Topical intranasal steroid + oral antibiotic vs. placebo + antibiotic or antibiotic alone OR: 0.72 (95% CI, 0.21 to 2.44)	NR
	1 study (N=15)	6 mos	Oral steroids + antibiotic vs. control + antibiotic OR: 0.15 (95% CI, 0.00 to 7.80)	NR
Williamson et al. 2010[56] Williamson et al. 2009[57]	NR	NR	Topical steroids vs. control Cure rate (A or C1 tympanogram in at least 1 ear) adjusted results (OR and RR) controlling for season, age, atrophy, and clinical severity score	NR
	G1: 96 G2: 98	1 mos	Diff in OR (adj): 0.934 (0.498 to 1.751) Diff in RR (adj): 0.97 (0.74 to 1.26)	NR
	G1: 86 G2: 86	3 mos	Diff in OR (adj): 1.451 (0.742 to 2.838) Diff in RR (adj): 1.23 (0.84 to 1.80)	Pass/Fail Criteria on sweep audiometry (fail at 2 or more frequencies at 25 dB in the better ear): G1: 52/83 (63%) G2: 47/81 (58%) Hearing loss from tympanograms, median (IQR) G1: 19.43 (14.64-1.21) G2: 21.15 (14.86-0.94) Baseline hearing G1: 30.97 (23.8-32.65) G2: 30.94(24.03-32.21)
	G1: 72 G2: 72	9 months	Diff in OR (adj): 0.822 (0.387 to 1.746) Diff in RR (adj): 0.90 (0.58 to 1.41)	Pass/Fail Criteria on sweep audiometry (fail at 2 or more frequencies at 25 dB in the better ear): G1: 44/74 (59%) G2: 34/67 (51%) Hearing loss from tympanograms, median (IQR) G1:19.56(14.88-0.84) G2: 17.89 (14.11-3.55)

Abbreviations: adj = adjusted; CI = confidence interval; dB = decibel; diff = difference; IQR = interquartile range; MA = meta-analysis; mos = months; OME = otitis media with effusion; OR = Odds ratio; RR = relative risk.

At 3 month followup, we found no differences in middle ear effusion, as measured by persisting OME, related to intranasal steroid use with adjunctive antibiotic treatment in both

arms, based on evidence from one trial that was included in the systematic review (N=59).[20] We also found no differences in cure rates based on evidence from the Williamson et al., study (N=172).[56] Similarly, at 6 month followup, persisting OME did not differ significantly between patients receiving oral steroid treatment plus antibiotic and controls plus antibiotic, based on evidence from one trial (N=15).[20] At 9 month followup, cure rates did not differ between topical steroids and control, based on the Williamson et al. study (N=144).[56]

Hearing Outcomes

Hearing did not differ between topical steroid and control groups, as measured at 3 and 9 months through audiometry and tympanometry based on one low risk of bias study (Table 26).[57] We did not find evidence related to hearing outcomes based on oral steroid treatment at 3 months or later.[20]

Other Outcomes

We found no evidence on other clinical outcomes, including OME recurrence, episodes of AOM, or vestibular function.

Nonpharmaceutical Interventions: Autoinflation

Description of Studies

The included evidence consisted of one recent Cochrane Collaboration review by Perera et al. (2009)[19] (Table 25) summarizing evidence from six RCTs of any form of autoinflation, a technique designed to increase oropharyngeal pressure via a nasal balloon or other process. Two different types of autoinflation devices were reviewed. One required the patient to actively inflate a balloon type device, whereas the other was a passive device in which the air was delivered into the nose while the patient swallowed. The review included five studies with children 3-12 years of age and one study of adults, 16-75 years of age. All studies were in comparison to no autoinflation. Other treatments (e.g., analgesics, antibiotics) were permitted as long as they were provided equally to both arms. The Cochrane review included one study with an end point of 2 weeks post-treatment, one study 4 weeks post treatment and one study approximately 2 months post treatment. The other three trials recorded outcomes at the end of treatment, the length of which differed, with no further outcome measurement. The systematic review was assessed as medium risk of bias.

Table 25. Characteristics of studies: Autoinflation

Study, Study Type, Country	Arm (N)	Diagnosis Criteria	Wait Period Between Diagnosis and Study	Inclusion/ Exclusion Criteria	Length of Study Followup	Age Range	Risk of Bias
Perera et al., 2009[19] Systematic review International	Autoinflation vs. control:6 trials (602 participants)	Tympanometry (type B or C2), either alone or in combination with simple or pneumatic otoscopy or audiometry.	Various	Include: RCTs; any form of autoinflation; other tx had to be given to both arms; OME diagnosis needed to include tympanometry	3 trials: at end of tx 1 trial: 4 wks post tx 1 trial: 2 mos 1 trial: 3 mos	3-12 yrs 4-11 yrs 6-75 yrs 3-10 yrs	Medium

Abbreviations: mos = months; OME = otitis media with effusion; RCT = Randomized controlled trial; tx = treatment; wks =

weeks; yrs = years

Key Points

- Relative to control groups, groups receiving autoinflation improved middle ear status as measured by tympanometry a month or less from treatment initiation based on two meta-analyses of two studies in one systematic review (low evidence) but groups did not differ in improvement in tympanometry at more than one month from treatment (insufficient evidence) (Table 27).
- Groups receiving autoinflation did not differ significantly from controls in measured hearing (PTA) at either the end of treatment or 4 weeks after treatment (insufficient evidence).
- No included studies reported on AOM, balance, or use of health care services (insufficient evidence).

Table 26. Strength of evidence for KQ 1: Autoinflation

Comparison (G1 vs. G2)	OME signs and symptoms	Objective hearing
Autoinflation vs. Control	**Low** Two MA (2; 185) Improvement in tympanogram with autoinflation at ≤ 1 mo, **Insufficient** One MA (2: 185) no difference in improvement in tympanogram at > 1 mo	**Insufficient** One MA (2: 125) No difference in HL improvement using PTA **Insufficient** One MA (2:179) No difference in average HL using PTA (4 wks post tx and end of tx)

Abbreviations: Mo = month; MA = meta-analysis; HL = hearing level; PTA = pure tone audiometry; tx = treatment

Detailed Synthesis

OME Signs and Symptoms Outcomes

The systematic review[19] presented results on improvement in middle ear status as measured with tympanometry (Table 27). Several of the trials reported improvement in tympanometric classification at different time points; in some studies, the outcomes were measured during the period of time that treatment was administered. The results were presented for different classifications at different time points. No study in the review reported on OME recurrence.

Table 27. Clinical outcomes: Autoinflation

Study	Arm (N)	Study Duration Until Outcome Measurement	Middle Ear Effusion	Measured Hearing
Perera et al., 2009[19]	MA: 3 studies (N=225) MA: 2 studies (N=185) MA: 2 studies (N=185)	≤ 1 month	Tympanometry improvement B or C2 to C1 or A: RR: 1.65 (95% CI 0.49, 5.56) B to C1 or A RR: 2.71 (95% CI 1.43, 5.12) C2 to C1 or A RR: 3.84 (95% CI 1.94, 7.59)	
	MA: 2 studies (N=185)	> 1 month	B or C2 to C1 or A: RR 1.89 (95% CI 0.77, 4,67)	
	MA: 2 studies (N=125)	End of treatment (3 weeks in 1 study and 3 months in the other study)		Improvement in HL ≥ 10 dB (measured by PTA RR 0.80 (95% CI 0.22, 2.88)
	MA 2 studies (N=179)	End of treatment in 1 study and 4 weeks after treatment in the other study		Average HL measured by PTA Weighted Mean Diff 7.02 (95% CI, -6.92, 20.96)

Abbreviations: dB = decibel; HL = hearing level; MA = meta-analysis; mo = month; PTA = pure tone average; RR = relative risk;

In one meta-analysis of three studies, the authors reported that, at 1 month or less, the autoinflation group did not have significant improvement from a B classification (a flat tracing usually indicative of the presence of middle ear fluid) at baseline, or C2 (highly negative curve, which is usually indicative of an abnormality) to a C1 classification (a moderately negative curve indicative as normal) or a tympanometric classification of A (considered to be normal).

Using data from two of the three trials included in the meta-analysis, the review authors reported two additional meta-analysis sub-analyses in which baseline tympanogram classifications were more narrowly combined. They found that autoinflation significantly improved middle ear status relative to no treatment in children with a baseline B classification (presence of middle ear fluid) and in children with a baseline C2 classification (negative pressure, indicative of abnormality) at followup of one month or less. However, in the meta-analysis of trials that examined ears at more than 1month from treatment initiation, they found no difference between autoinflation patients and controls in rates of improvement in tympanometry (i.e., from B or C2 classifications indicating presence of fluid or an abnormal to C1 or A classifications, indicating as normal middle ear status).

Hearing Outcomes

Two meta-analyses examined hearing outcomes. The autoinflation and control groups did not differ in hearing outcomes at the end of treatment of varying lengths or at 4 weeks after treatment, measured either by a change in hearing level as measured by pure tone audiometry as a discrete outcome or by the PTA hearing threshold.

Other Outcomes

Other relevant outcomes, such as episodes of AOM, OME recurrence, or vestibular function were not discussed as a function of treatment.

KQ 2: Comparative Effectiveness of Interventions: Functional and Quality of Life Outcomes

Surgical Interventions: Tympanostomy Tube Comparisons

No studies reported on functional or quality-of-life outcomes.

Surgical Interventions: Tympanostomy Tubes Versus Watchful Waiting or Myringotomy

Key Points

- Speech and language development did not significantly differ between children receiving tympanostomy tubes and those receiving watchful waiting after 6 to 9 months and cognitive development after less than and more than a year (Table 28). Strength of evidence is low.
- Evidence is insufficient on differences in behavioral outcomes (mixed) and quality-of-life outcomes (one study) comparing tympanostomy tube and watchful waiting groups at less than one year. At one year or more, two studies found no difference in behavior (strength of evidence was low).
- No evidence was found comparing tympanostomy tubes and myringotomy in relation to other functional outcomes. Strength of evidence is insufficient.

Table 28. Strength of evidence for KQ 2: Tympanostomy tubes versus watchful waiting or myringotomy

Comparison (G1 vs. G2)	Speech/Language and Cognitive Development	Behavior	Quality of Life
TT vs. watchful waiting 6 to 9 months	**Low** No difference MA: 3, 394; Study: 1, 160	**Insufficient** Mixed results Studies: 2, 358	**Insufficient** No difference Study: 1, 176
TT vs. watchful waiting 1 year or more	**Low** No difference Study: 1, 393	**Low** No difference Studies: 2, 347	**Insufficient** No difference Study: 1 Study: 1, 176
TT vs. Myringotomy	**Insufficient** No studies	**Insufficient** No studies	**Insufficient** No studies

Abbreviations: TT = tympanostomy tubes

Detailed Synthesis

Speech/Language and Cognitive Outcomes

Two systematic reviews evaluated functional and quality of life outcomes in subjects who received tympanostomy tubes compared with watchful waiting or myringotomy (Table 29).[15, 17] Browning et al., (2010) performed meta-analysis of three studies measuring differences in language comprehension and language expression at six to nine months between tympanostomy tube and watchful waiting groups and found no significant difference (mean difference=0.09 [95% CI, -0.21 to 0.39] and mean difference=0.03 [95% CI, -0.41 to 0.49], respectively). Browning et al. identified a single study that evaluated cognitive development at 9 month followup and found no significant difference.[17]

Table 29. Functional and health-related quality-of-life outcomes: Tympanostomy tubes versus other treatments

Study	Arm (N)	Study Duration Until Outcome Measurement	Speech/Language Cognitive Development	Behavior	Quality of Life
Browning et al., 2010[17]	MA=3 studies (N=394)	6-9 mos	Language Comprehension, Bil TT vs. WW: Mean Difference 0.09 (95% CI, -0.21, 0.39)	NR	NR
	MA=3 studies (N=393)	6-9 mos	Language Expression TT vs. WW: Mean Difference 0.03 (95% CI, -0.42, 0.49)	NR	NR
	1 study (N=160)	9 mos	Griffiths Mental Development Mean Cognitive Index TT vs. WW 106.5 vs. 104.2 (95% CI, -2·58 to 7·04) (p=ns)	NR	NR
	1 study (N=393)	3 yrs	McCarthy Mental Development Mean General Cognitive Index TT vs. WW 99 vs. 101 (95% CI, -4.1 to 1.1)	Child Behavior Checklist Mean Total Problem Score TT vs. WW 50 vs. 49 (95% CI, -0.6 to 3.4)	NR
	1 study (N=176)	6 mos	NR	The Erickson child Mean scores TT vs. WW Affection 4.4 vs. 4.6 Avoidance 6.3 vs. 6.5 Compliance 5.1 vs. 5.2 Negativism 6.6 vs. 6.7 Reliance 6.5 vs. 6.7 MANOVA Hotelling Trace p=0.19	The TAIQOL Mean scores TT vs. WW Vitality 3.3 vs. 3.3 Appetite 5.0 vs. 4.7 Communication G1: 6.7 vs. 5.8 Motoric 4.4 vs. 4.4 Social 3.5 vs. 3.5 Anxiety 4.3 vs. 4.1 Aggression 11.9 vs. 11.1 Eating 3.3 vs. 3.5 Sleeping 6.8 vs. 6.6 MANOVA Hotelling Trace p= 0.22

Table 29. Functional and health-related quality-of-life outcomes: Tympanostomy tubes versus other treatments (continued)

Study	Arm (N)	Study Duration Until Outcome Measurement	Speech/ Language Cognitive Development	Behavior	Quality of Life
Browning et al., 2010[17] (continued)	(N=165)	12 mos	NR	Affection 4.5 vs. 4.9 Avoidance 6.5 vs. 6.9 Compliance 5.2 vs. 5.6 Negativism 6.6 vs. 6.9 Reliance 6.6 vs. 6.8 MANOVA Hotelling Trace p=0.38	Vitality 3.1 vs.3.2 Appetite 5.3 vs.4.9 Communication 5.9 vs.5.6 Motoric 4.2 vs.4.2 Social 3.5 vs. 3.5 Anxiety 4.6 vs. 4.3 Aggression 11.8 vs.11.5 Eating 3.3 vs. 3.4 Sleeping 6.4 vs. 6.4 MANOVA Hotelling Trace p=0.94
Hellstrom et al., 2011[15]	1 study (n=182)	9 mos	NR	Richman Behavioral Scale, % with Problems TT vs. WW 30% vs. 47% (95% CI, -33% to –2%) p=0.031 (favors tx)	.
		18 mos	NR	24% vs. 20% (95% CI, -10% to 19%) p=0.66	

Abbreviations: bil = bilateral; MA = meta-analysis; mos = months; TAIQOL = TNO-AZL Infant Quality of Life; TT = tympanostomy tubes; tx = treatment; WW = watchful waiting; yrs = years; vs.=versus

Behavior Outcomes

Both Browning et al. and Hellstrom et al. identified studies that examined differences in behavioral outcomes in studies of children with OME receiving treatment with tympanostomy tubes or watchful waiting.[15, 17] Results were mixed. Erickson Child mean scores did not significantly differ at 6 and 12 month followup[17], nor did Child Behavior Checklist Problem Scores at 3 year followup.[17] However, Richman Behavioral Scale scores were better in the tympanostomy tube group at 9, but not 18 month followup.[15]

Quality of Life

Browning et al. found no significant differences in a single study that evaluated quality of life using the TAIQOL (a generic quality-of-life measure for 1to 4 year olds) in children with OME who received tympanostomy tubes or watchful waiting at 6 and 12 month followup.[17]

Surgical Interventions: Tympanostomy Tubes Plus Adenoidectomy Versus Myringotomy Plus Adenoidectomy or Adenoidectomy Alone

Key Findings

- One small study examined differences in quality-of-life between children receiving adenoidectomy and tympanostomy tubes and those receiving adenoidectomy and myringotomy. Strength of evidence was insufficient.
- Strength of evidence was insufficient for all speech/language, cognitive and behavioral outcomes because these outcomes were not examined.

Detailed Synthesis

Only one small study (N=52) evaluated differences quality-of-life outcomes between children receiving tympanostomy tubes and adenoidectomy and those receiving myringotomy and adenoidectomy [52] (Table 30). While only one group improved from baseline, the difference between them was not significant.[52]

Table 30. Functional and health-related quality-of-life outcomes: Tympanostomy tubes plus adenoidectomy versus myringotomy plus adenoidectomy or adenoidectomy alone

Study	Arm (N)	Study Duration Until Outcome Measurement	Speech/ Language Cognitive Development	Quality of Life
Vlastos et al., 2011[52]	G1: TT + adenoidectomy (N=25) G2: Myringotomy + adenoidectomy (N=27)	6 months		OM-6 Score Score G1: 1.88 G2: 2.04 Mean Difference: -.0.16 (95% CI, -0.43 to 0.10) Change from Baseline G1: -0.38 G2: -0.00 mean change: -0.38 (95% CI, -0.65 to -0.10)
		12 months		Score G1: 1.84 G2: 2.04 Mean Difference: -0.20 (95% CI, -.0.57 to 0.17) Change from Baseline G1: -0.32 G2: 0.01 mean change: -0.23 (95% CI, -0.76 to 0.11)

Abbreviations: OM-6 = Otitis Media -6

Surgical Interventions: Myringotomy Comparisons

Key Findings

- We found no evidence examining functional outcomes or health related quality of life.

Surgical Interventions: Myringotomy Plus Adenoidectomy Comparisons

Key Findings

- We found no evidence examining functional outcomes or health related quality of life.

Surgical Interventions: Adenoidectomy Versus Other Interventions

Key Findings

- We found no evidence examining functional outcomes or health related quality of life.

Pharmaceutical Interventions: Oral or Topical Nasal Steroids

Key Points

- Patients receiving topical intranasal steroids versus controls did not differ on the quality-of-life outcome of reported hearing at 3 months (insufficient evidence, one small study).
- Patients receiving topical intranasal steroids versus controls, both receiving antibiotics, did not differ at 3 months on a quality-of-life symptom score (low strength of evidence, one study).
- No study reported on oral steroids versus controls (with or without antibiotics) or oral steroids plus antibiotics versus controls (insufficient evidence).
- No study reported on speech or language outcomes, cognitive development, or behavioral outcomes (insufficient evidence).

Detailed Synthesis

Topical nasal steroids did not differ from controls in relation to symptom scores (one study from the systematic review [N=39])[20] (Table 31). One other study did not find significant differences between groups in parent-reported hearing difficulties or days with hearing loss.[57]

Table 31. Functional and health-related quality-of-life outcomes: Oral or topical nasal steroids

Study	Arm (N)	Study Duration until outcome measurement	Quality of Life
Thomas et al., 2010[20]	1 study (N=39)	3 mos	Symptom score Topical intranasal steroid + oral antibiotic vs. control + antibiotic or antibiotic alone Peto odds ratio: -4.50 (95% CI, -10.28 to 1.28)
Williamson et al., 2010[56] Williamson et al., 2009[57]	Baseline: (N=196)	Baseline: G1: 6.06 (2.83-8.57) G2: 5.88 (2.33-7.60)	Parent-reported hearing difficulties, median(IQR) Topical intranasal steroid vs. control
	G1: 86 G2: 86	3 mos	G1: 5.54 (0.90-8.43) G2: 3.92 (0.90-7.60) (p=NS)[a]
	G1: 72 G2: 72	9 mos	G1: 2.33 (0.21 to 7.60) G2: 2.33 (0.42-6.60) (p=NS)[a]
	G1: 86 G2: 86	3 mos	Days with hearing loss, median (IQR) G1: 4 (0 to 24.5) G2: 4 (0 to 18.5) p=0.45

[a]Calculated by authors

Abbreviations: CI = confidence interval; IQR = interquartile range; mos = months; N = number; NS = not sufficient.

Nonpharmaceutical Interventions: Autoinflation

Key Findings

- We found no evidence examining functional outcomes or health related quality of life.

KQ 3: Harms or Tolerability

Surgical Interventions: Tympanostomy Tube Comparisons

Key Points

- Otorrhea rates differed by tube type; Paparella tubes had the highest rates reported in two observational studies (insufficient strength of evidence).
- For other side effects, such as perforation, tympanosclerosis, atrophy, cholesteatoma, or granulation, evidence was either not available at all (no studies) or studies provided either conflicting results or no difference (insufficient evidence).

Detailed Synthesis

The systematic review by Hellstrom et al.[15] identified nine studies that compared type of tympanostomy tube or insertion technique and various side effects (Table 32).[15] All seven newly identified studies comparing the effectiveness of tubes examined various potential harms as well.[35-41]

Table 32. Treatment harms or tolerability: Tympanostomy tube comparison studies

Study	Arm (N)	Study Duration Until Outcome Measurement	Repeat Tubes	Otorrhea	Perforation	Cholesteatoma	Tympano-sclerosis	Occlusion	Granulation
Wielinga and Smyth, 1990[36]	G1: Silicon tube (N=15) G2: Teflon Armstrong tube (N=15)	NS	G1: 20% G2: 47%	G1: 13% G2: 20%	G1: 6% G2: 6%	G1: 0% G2: 0%	NR	NR	NR
Abdullah, Pringle and Shah, 1994[40]	G1: Trimmed high grade silicone Shah permavent tube (N=25) G2: Polyethylene Shah tube (N=25)	29 months	NR	NR	NR	NR	G1: 41% G2: 65% p=NR	NR	NR
Licameli et al., 2008[39]	G1: Phosphorylcholine -coated fluoroplastic Armstrong tube (N=70) G2: Uncoated fluoroplastic Armstrong tube (N=70)	24 months	NR	G1: 8.7% G2: 7.5% p=0.74	G1: 4.0% G2: 0% p=0.24	NR	NR	G1: 10.3% G2: 13.4 p=0.53	G1: 4.4% G2: 6.0% p=0.66
Iwaki et al., 1998[37]	G1: Teflon Shepard tube (N=75) G2: Silicone Goode-T tube (N=39) G3: Silicone Paparella II tube (N=106)	24 months	NR	G1: 12% G2: 36% G3: 38% p<0.01	G1: 0% G2: 7.7% G3: 10.4% p<0.05	G1: 1.3% G2: 0% G3: 0% p=NR	NR	NR	G1: 0% G2: 0% G3: 7.5% p<0.05

Table 32. Treatment harms or tolerability: Tympanostomy tube comparison studies (continued)

Study	Arm (N)	Study Duration Until Outcome Measurement	Repeat Tubes	Otorrhea	Perforation	Cholesteatoma	Tympano-sclerosis	Occlusion	Granulation
Slack, Gardner, and Chatfield, 1987[38]	G1: Shepard tube (N=214) G2: Shah tube (N=70) G3: Paparella tube (N=275) G4: Goode tube (N=4) G5: Reuter Bobbin (N=28) G6: Other types (N=63)	6 months		G1: 6.2% G2: - G3: 0%	NR	NR		NR	
		12 months		G1: 6.5% G2: 5.9% G3: 20% p>0.05					
		18 months		G1: 3.5% G2: 0% G3: 37.5% p<0.001					
		24 months		G1: 6.2% G2: 13.0% G3: 38.4% p<0.001					
		30 months		G1: - G2: 0% G3: 53.6% p<0.01					
Oversen et al., 2000[35] Demark	G1: TT + N-acetylcysteine instilled (N=37) G2: TT + placebo vehicle (N=38) G3: TT (N=75)	29 months		G1: 16% G2: 39% G3: 69% p<0.025	G1: 24% G2: 13% G3: 16% p>0.15				

Table 32. Treatment harms or tolerability: Tympanostomy tube comparison studies (continued)

Study	Arm (N)	Study Duration Until Outcome Measurement	Repeat Tubes	Otorrhea	Perforation	Cholesteatoma	Tympano-sclerosis	Occlusion	Granulation
McRae et al., 1989[41]	G1: Shah Teflon tube + aspiration prior to placement (N=55) G2: Shah Teflon tube (no aspiration) (N=55)						G1: 66% G2: 47% p<0.05		
Hellstrom et al., 2011[15]	24 Studies included 5475 subjects			3 Study: Touch vs. non-touch during surgery, no difference in otorrhea. 2 studies: Otic drops at insertion decreased otorrhea In both			1 Study: Shah worse than Mini-Shah after 2 years, p<0.001		

Otorrhea

Otorrhea rates were found to vary by type of tympanostomy tube inserted. Based on observational evidence, Paparella tympanostomy tubes had significantly higher otorrhea rates than Shepard and Shah tubes at 12, 24, and 30 months after placement in one study[38] and at 24 months in comparison to Teflon Shepard tubes in a second.[37]

Otorrhea rates were not significantly related to infusion of N-acetylcysteine at the time of insertion.[35] The systematic review included two studies that reported that topical antibiotics at the time of tympanostomy tube insertion decreased rates of otorrhea (data not provided).[15]

Repeat Tube Placement

In one trial, patients who received Teflon Armstrong tubes were more likely to undergo repeat tube placement than those receiving the silicone Goode-T tube (47% vs. 20%, respectively; no p value reported).[36] In one study, N-acetylcysteine infusion at the time of placement significantly reduced the need for repeat tube placement thereafter ($p<0.025$).[35]

Other Harms

Other side effects and potential harms found in studies included risk of cholesteatoma, occlusion, rate of tympanosclerosis, and presence of granulation tissue. Groups did not differ significantly in either cholesteatoma formation by tube type,[36, 37] or occlusion rates.[39] One study in the systematic review examined tympanosclerosis as an undesired outcome.[58] Standard Shah tubes had higher tympanosclerosis rates than mini-Shah tubes, but the standard Shah tubes also had higher retention rates. Abdullah et al. found a possible increased rate of tympanosclerosis for polyethylene Shah tubes compared with silicone permavent Shah tube (65% vs. 41%, no p value reported). Aspiration prior to tube placement increased the tympanosclerosis rate ($p<0.05$).[41] Finally, Iwaki et al. demonstrated higher rates of granulation tissue at 24 months for silicone Paparella tube compared to either Goode-T silicone or Teflon Shepard tubes (7.5%, 0%, 0% respectively, $p<0.05$).[37]

Surgical Interventions: Tympanostomy Tubes Versus Watchful Waiting or Myringotomy

Key Points

- Otorrhea and tympanosclerosis were found to occur more frequently in ears that had tympanostomy tubes placed previously (Strength of evidence low).
- Evidence was insufficient for all other side effects or harms.

Detailed Synthesis

Otorrhea/AOM/Otalgia

Both identified studies[44, 45] and systematic reviews[15, 17] evaluated ear related side effects (Table 33). Otorrhea occurs with a perforated tympanic membrane or an in-place tympanostomy tube, so that outcome is unlikely to occur with watchful waiting.

Table 33. Harms or tolerability: Tympanostomy tubes versus watchful waiting or myringotomy

Study	Arm (N)	Study Duration Until Outcome Measurement	Repeat Tubes/ Treatment Failure	Otorrhea/ AOM/ Otalgia	Atrophy	Tympano-sclerosis Myringo-sclerosis	Perforation	Choleste-atoma	Granulation
Browning et al., 2010[17]	1 study by ear (N=78)	1 year	NR	NR	NR	TT vs. WW 38% vs. 1%	NR	NR	NR
	1 study by child (N=248)	24 months	NR	Otorrhea 2%	NR	TT vs. WW 27% vs. 0%	<1%	NR	NR
	1 study (N=187)	6 months	NR	Otorrhea TT vs. WW 49% (95% CI, 39% to 60%) vs. 10% (95% CI, 4% to 16%)	NR	NR	NR	NR	NR
	1 study (N=236)	NR	NR	AOM TT vs. non-tubed 27% vs. 11%	NR	NR	NR	NR	NR

Table 33. Harms or tolerability: Tympanostomy tubes versus watchful waiting or myringotomy (continued)

Study	Arm (N)	Study Duration Until Outcome Measurement	Repeat Tubes/ Treatment Failure	Otorrhea/ AOM/ Otalgia	Atrophy	Tympano-sclerosis Myringo-sclerosis	Perforation	Cholesteatoma	Granulation
Hellstrom et al., 2011[15]	1 study (N=277)						TT vs. M 2.4% vs. 3%		
	1 study (N=111)						TT vs. M vs. no tx TT 5.6%	TT vs. M vs. no tx No surgery 5%	
	1 study (N=429)	5,6 years of age			TT vs. control TT worse RR 17.4	TT vs. control TT worse RR diff: 24.5			
	1 study (N=224)				TT vs. M No difference	TT vs. M TT worse, (p<.001)			
	1 study (N=150)				TT vs. control TT worse 13% vs. 1.3%	TT vs. control TT worse 33% vs. 6.7%			TT vs. control TT worse 4% vs. 0%
	1 study (N=222)	5 years 10 years			TT vs. control TT worse RR 80% RR 80%				
Koopman et al., 2004[45]	G1: TT + cold knife Myringotomy (N=208) G2: Laser Myringotomy (N=208)		NR	Otorrhea G1 more often than G2 p=0.0020 Otalgia without inflammation G1: 1 G2: 0	NR	NR	NR	NR	NR	NR

Table 33. Harms or tolerability: Tympanostomy tubes versus watchful waiting or myringotomy (continued)

Study	Arm (N)	Study Duration Until Outcome Measurement	Repeat Tubes/ Treatment Failure	Otorrhea/ AOM/ Otalgia	Atrophy	Tympano-sclerosis Myringo-sclerosis	Perforation	Cholesteatoma	Granulation
Mandel et al., 1989[44]	Without "significant" hearing loss" G1: Myringotomy G2: Myringotomy + TT G3: No surgery with significant hearing loss G4: Myringotomy G5: Myringotomy + TT		Tx failure: G1: 0.53 G2: 0 G3: 0.59 G4: 0.75 G5: 0 p=NS	Otorrhea episodes/ person yr G1: 0.15 G2: 0.41 G3: 0.23 G4: 0.34 G5: 0.61 In non-TT groups this is tx failures received tubes	NR	NR	NR	G3: 1 ear	NR

Abbreviations: AOM = acute otitis media; M = myringotomy; RR = relative risk; TT = tympanostomy tubes; tx = treatment; WW = watchful waiting; yr = year;

Tympanosclerosis

The Hellstrom et al. and Browning et al. systematic reviews both considered rates of tympanosclerosis in tympanostomy tube versus watchful waiting or myringotomy groups.[15, 17] Tympanosclerosis rates were higher in the tympanostomy tube groups in five studies included in the two reviews in subsequent examinations after the tubes had been extruded.[15, 17]

Atrophy

Hellstrom identified four studies that evaluated atrophy subsequent to tympanostomy tubes versus myringotomy or watchful waiting.[15] Results were mixed. Tympanostomy tubes were associated with higher rates of atrophy in two studies and no different in the other two studies.

Other Harms

Hellstrom identified two studies[15] and Browning[17] identified one study that evaluated perforation following tympanostomy tube insertion in comparison with myringotomy or no treatment. In all studies, typanostomy tubes were associated with low rates of perforation, but no significance tests were reported. Similarly, cholesteatoma was reported in one study in Hellstrom; a low rate was only reported in the no treatment group, but no significance tests were reported. Hellstrom reported on granulation in one study comparing tympanostomy tubes with control. Tympanostomy tubes were associated with higher rates of granulation (4% to 0%), but no significance tests were reported.

Surgical Interventions: Tympanostomy Tubes Plus Adenoidectomy Versus Myringotomy Plus Adenoidectomy or Adenoidectomy Alone

Key Points

- The risk of tympanosclerosis were higher in the tympanostomy tube group than in groups either with no treatment in addition to adenoidectomy or myringotomy in addition to adenoidectomy (strength of evidence was low).
- Evidence for all other harms or side effects was insufficient because evidence was sparse or results were mixed (strength of evidence insufficient for all outcomes).

Detailed Synthesis

Tympanosclerosis

Two studies comparing tympanostomy tubes with myringotomy plus adenoidectomy or adenoidectomy alone found higher rates of tympanosclerosis in the tympanostomy tube group (Table 34).[46, 48, 53]

Otorrhea

Three studies that evaluated otorrhea after tympanostomy tube placement or myringotomy found mixed results.[49-51] One found no difference between interventions.[49] A second found higher rates with tympanostomy tubes.[50] A third found a lower rate in the tympanostomy tube group.[52]

Table 34. Harms or tolerability: Tympanostomy tubes plus adenoidectomy versus myringotomy plus adenoidectomy or adenoidectomy alone

Study	Arm (N)	Study Duration Until Outcome Measurement	Repeat Tubes/ Treat-ment Failure	Otorrhea	Perforation	Tympano-sclerosis Myringo-sclerosis
Brown et al., 1978[46]	G1: TT + AD (N=55) G2: AD (N=55)					G1: N=23 G2: N=0 p<0.05
Lindholdt, 1977[48]	G1: TT+ AD (N=91 ears) G2: AD (N=91 ears)		G1: 13 G2: 6			
D'Eredita and Shah, 2006[49]	G1: CDLM + AD (N=15) G2: TT + AD (N=15)	30 days		Otorrhea G2: 4 reports		
		2 months		G1: 2 reports		
		3 months		G2: 4 reports		
		1 year			G1: 0 G2: 1	
Popova et al., 2010[50]	G1: TT + AD (N=42) G2: M (N=36)			Otorrhea G1: 40% G2: 0%		
Shishegar and Hobhoghi, 2007[51]	G1: TT + AD (n=30 ears) G2: M + AD (N=30 ears)	>6 months		Otorrhea G1: 7% G2: 27% Over 6 mos. Percentage of TT occluded,: 17%		
Vlastos et al., 2011[52]	G1: TT + AD (N=25) G2: M + AD (N=27)		G2: 20% (tubes in non tube group)			
Tos and Stangerup, 1989[53]	G1: TT + AD (N=146) G2: M+ AD (N=146)					G1: 59% G2: 13% p<0.05

Abbreviations: AD = adenoidectomy; M = myringotomy; mos = months; TT = tympanostomy tubes, CDLM = Contact diode laser myringotomy

Surgical Interventions: Myringotomy Comparisons

Key Points

- We found no evidence examining harms or tolerability.

Surgical Interventions: Myringotomy Plus Adenoidectomy Comparisons

Key Points

- We found no evidence examining harms or tolerability.

Surgical Interventions: Adenoidectomy Versus Other Interventions

Key Points

- We found no evidence examining harms or tolerability.

Pharmaceutical Interventions: Oral or Topical Nasal Steroids

Key Points

- Groups did not differ significantly in mild adverse events such as nasal stinging (low strength of evidence).
- Evidence concerning serious harms was sparse (insufficient strength of evidence)

Detailed Synthesis

The systematic review focusing on steroid treatment for OME found no serious or lasting harms reported in five studies of oral steroids and two studies of topical steroids (Table 35).[20] The Williamson et al. study found no significant difference at 3 months between the topical steroid group and control in relation to stinging nose, nose bleed, dry throat, or cough.[56, 57]

Table 35. Treatment harms or tolerability: Oral or topical nasal steroids

Study	Arm (N)	Study Duration Until Outcome Measurement	Serious Or Lasting Harm Outcomes	Mild Adverse Outcomes
Thomas et al., 2010[20] Systematic review International	NR	Variable	No serious or lasting harms reported in 5 studies of oral steroids or 2 studies of topical intranasal steroids.	Some studies mentioned mild adverse outcomes: vomiting, diarrhea, dermatitis, transient nasal stinging, and epistaxis.
Williamson et al. 2010[56] Williamson et al. 2009[57] RCT UK	G1: 9/85 (11%) G2: 9/85 (11%) G1: 10/86 (12%) G2: 6/84 (7%) G1: 10/85 (12%) G2: 7/83 (8%) G1: 19/86 (22%) G2: 11/83 (13%)	3 months		Stinging nose: RR: 1.00 (95% CI, 0.42 to 2.40 Nose bleed: RR: 1.63 (95% CI, 0.62 to 4.28) Dry throat: 1.40 (95% CI, 0.56 to 3.49) Cough: 1.67 (95% CI, 0.85 to 3.29)

Abbreviations: RR = relative risk

Nonpharmaceutical: Autoinflation

Key Points

- No quantitative information on rates of serious or mild harms was provided (Insufficient strength of evidence).

Detailed Synthesis

The systematic review stated that there were no serious or lasing harms reported in the six studies of autoinflation but no data was provided (Table 36).[19] It reported that, in one trial, a patient stopped treatment because of pain.[19]

Table 36. Treatment harms or tolerability: Autoinflation

Study	Arm (N)	Study Duration Until Outcome Measurement	Serious Or Lasting Harms	Mild Adverse Outcomes
Perera et al., 2009[19]	NR	Variable	"None of the studies demonstrated a significant difference in the incidence of side effects between the control or intervention groups"	One trial in the systematic review reported that "that one patient stopped the treatment due to the pain caused by the procedure"

KQ 4: Comparative Effectiveness of Interventions for Subgroups of Patients

Surgical Interventions: Tympanostomy Tube Comparisons

Key Points

- No studies reported on patient subgroups. Evidence is insufficient.

Surgical Interventions: Tympanostomy Tubes Versus Watchful Waiting or Myringotomy

Key Points

- No studies reported on patient subgroups. Strength of evidence insufficient.

Surgical Interventions: Tympanostomy Tubes Plus Adenoidectomy Versus Myringotomy Plus Adenoidectomy or Adenoidectomy Alone

Key Points

- A single study evaluated subgroups of patients (strength of evidence was insufficient).

Detailed Synthesis

One study of children with sleep apnea and OME[52] did not find important differences in hearing thresholds after placement of tympanostomy tubes or myringotomy See Tables 14 and 30). Among this group of children who had adenoidectomies for sleep apnea and also had OME,

quality of life did not change at 12 months between tube or myringotomy groups. At 6 months, results were mixed with some measures improving more quickly in the tympanostomy tube group, while on other measures tube and myringotomy were the same.[52] Twenty percent of the children who initially received a myringotomy, eventually also received tympanostomy tubes.[52] We were unable to compare this rate to other identified studies.

Surgical Interventions: Myringotomy Comparisons

Key Points

- The one study examining differences in myringotomy procedures did not examine the comparative effectiveness of the two approaches within subgroups of patients. Evidence is insufficient.

Surgical Interventions: Myringotomy With Adenoidectomy Comparisons

Key Points

- No studies of these interventions reported on patient subgroups. Evidence is insufficient.

Surgical Interventions: Adenoidectomy Versus Other Interventions

Key Points

- No studies of these interventions reported on patient subgroups. Evidence is insufficient.

Pharmaceutical Interventions: Oral or Topical Nasal Steroids

Key Points

- No studies of these interventions reported on patient subgroups. Evidence is insufficient.

Nonpharmaceutical Interventions: Autoinflation

Key Points

- A subgroup of adults who received autoinflation had better middle ear effusion outcomes than controls at end of treatment and 50 days after treatment (Low strength of evidence, one study).

Detailed Synthesis

One study in the Cochrane review on autoinflation[19] included adults (Table 37).[59] The treatment intervention was a BD Politzer device used twice a day for 10 days, with or without antibiotics. The control group received equal care except for the intervention. Followup and adherence were 100 percent. The outcome measure was a composite measure of recovery from OME based on pneumo-otoscopy, tympanometry, and audiometry. Individuals in the autoinflation group were significantly more likely to experience a complete recovery than those in the control group at both the end of treatment ($p<0.001$) and at 50 days after treatment

(p<0.001). Similarly, the ears of the participants receiving autoinflation had better recovery rates than control ears at both time points (p<0.001).

Table 37. Comparative effectiveness for adults: Autoinflation

Study	Arm (N)	Study Duration until outcome measurement	Composite measure of recovery
Perera et al., 2009[19]	Autoinflation vs. control (n=198)	End of treatment:(10 weeks)	Individuals: 50.6% vs. 3.8% Ears: 49.2% vs. 3.9% (p<0.001) (favors autoinflation)
		50 days after treatment	Individuals: 55.2% vs. 11% Ears: 57.8% vs. 11.8% (p<0.001) (favors autoinflation)

KQ 5: Comparative Effectiveness by Health Care Factors

Key Points

- No included studies for any intervention comparisons examined effectiveness by any health care factors.

Discussion

Key Findings and Strength of Evidence

This systematic review addressed the comparative effectiveness of treatments for otitis media with effusion (OME). OME is characterized by Eustachian tube dysfunction, the accumulation of fluid in the middle ear and most commonly is a condition that affects children. Health care providers have been particularly concerned when fluid persists for a relatively long period of time (e.g., 3 months or more) and if it reduces hearing because it may result in functional limitations and have long term sequelae.

Various approaches have been studied for treating OME. Sometimes investigators used a single treatment alone; sometimes they combined two or more approaches. In this review, we focused on the following interventions and comparisons among them: surgical (tympanostomy tubes, myringotomy, and adenoidectomy); nonpharmacological interventions (autoinflation); pharmacological (oral or nasal steroids); complementary and alternative medicine (CAM); and other treatment strategies (watchful waiting). The effectiveness of these interventions has been generally studied in pediatric samples that included a wide range of ages. The focus of this review was to compare the relative benefits and harms of these treatment approaches overall and then specifically in particular subpopulations of interest who may be particularly affected by OME (e.g. children with preexisting hearing limitations, craniofacial abnormalities or Down syndrome) or for whom little is known (adults). As discussed in the introduction, we did not consider hearing aids, antihistamines and decongestants, or antibiotics.

Overview

Overall, the evidence included 13 randomized controlled trials (RCTs), four nonrandomized trials (e.g., studies comparing left and right ears), three observational studies, and five systematic reviews (four of which were limited to RCTs). By treatment comparison, the literature included:

- Surgical approaches:
  - Tympanostomy tubes compared by type of tube or procedure approach
  - Tympanostomy tubes versus myringotomy or nonsurgical interventions (delayed treatment or watchful waiting),
  - Tympanostomy tubes plus adenoidectomy versus myringotomy plus adenoidectomy
  - Tympanostomy tubes plus adenoidectomy versus adenoidectomy,
  - Myringotomy versus myringotomy, comparison of different approaches, various combinations of myringotomy plus adenoidectomy; and
  - Adenoidectomy versus nonsurgical interventions or myringotomy.
- Pharmacological interventions, specifically oral and topical nasal steroids, and
- Nonpharmacological interventions, specifically, autoinflation.

We had no studies meeting inclusion criteria on any CAM interventions; all watchful waiting studies were combined with myringotomy so we have no independent body of information about other interventions.

Notably, we restricted our review to treatments for OME. Although clinicians use many of these treatments for patients with recurrent acute otitis media (AOM), we included only studies

from which we could obtain evidence for pure OME populations. We did not restrict concomitant conditions as long as the participants had OME. Although we had hoped to be able to provide evidence for different subpopulations, the review pertains mainly to typically developing children across a wide age range. We were unable to find studies on individuals with cleft palate or sensorineural hearing loss that met our inclusion criteria, and we found only one study that targeted individuals 16 to 75 years of age.

We tried to examine a broad range of clinical, functional, and quality-of-life outcomes and harms of treatment. Although most of the studies examined middle ear status (e.g., presence of effusion or recurrence of OME), and many examined hearing and harms of treatment, only a handful included measures of speech, language, behavior, or quality of life. None of the studies examined vestibular function or health care utilization. Thus, our statements about evidence are limited primarily to middle ear status, hearing, and harms.

We summarize the strength of evidence for interventions, comparisons, and outcomes on which we had studies of at least low or medium risk of bias. Strength of evidence grades are developed from ratings on four domains: overall risk of bias, directness of the evidence or the comparisons, consistency, and precision of estimates.[30] We did not evaluate other strength of evidence domains (e.g., publication bias, dose-response relationships). Strength of evidence can have one of four grades—high, moderate, low, or insufficient; insufficient evidence arises when we had no studies at all addressing the particular topic, when we had only a single small study, when available studies were sufficiently inconsistent, indirect, or imprecise as to preclude drawing any conclusions or when differences in treatments appear to show no difference among studies that may be underpowered or clinical thresholds for minimal differences have not been established.

KQ 1: Clinical Outcomes

For this key question (KQ), we sought evidence on the effectiveness of surgical and other interventions on a range of clinical outcomes including recurrent middle ear effusion, recurrent AOM, and measured hearing; as noted, we had no studies that reported on vestibular function or use of health care services.

Summary of Findings and Strength of Evidence

The evidence base consisted of 20 studies reported in 21 articles (13 RCTs, 3 NRCTs and 4 observational studies) and 5 systematic reviews that provided information on clinical outcomes for various surgical interventions, autoinflation, or steroids.

Table 38 summarizes the OME interventions on which we had low, moderate or high strength of evidence for clinical outcomes.

Table 38. Strength of evidence for interventions to improve clinical outcomes

Intervention and Comparator	Number of Studies (Sample sizes)	Outcome and Results	Strength of Evidence
Tympanostomy tubes vs. watchful waiting or myringotomy	MA of 3 RCTs (N=574)	Tympanostomy tubes decreased persistent middle ear effusion at 1 year: 32% less time (95% CI, 17% to 48%)	High for benefit
	MA of 3 RCTs (N=426)	Tympanostomy tubes decreased persistent middle ear effusion at 2 years: 13% less time (95% CI, 8% to 17%)	Moderate for benefit
	MA of 3 RCTs (by ears) (N=230)	Tympanostomy tube groups had better measured hearing at 4-6 months: -10dB (95% CI, -19.12 to -1.05)	High for benefit
	MA of 3 RCTs (N=523)	Tympanostomy tube groups had better measured hearing at 6-9 months: -4.20dB (95% CI, -4.00 to -2.39)	Moderate for benefit
	MA of 3 RCTs (by ears) (N=234)	No difference between groups in measured hearing at 7-12 months: -5.18dB (95% CI, -10.43 to 0.07)	Low for no difference
	MA of 2 RCTs (N=328)	No difference between groups in measured hearing at 12 months: -0.41dB (95% CI, -2.37 to 1.54)	Moderate for no difference
	MA of 2 RCTs (N=283)	No difference in measured hearing between groups at 18 months: -0.02 dB (95% CI, -3.22 to 3.18)	Moderate for no difference
Tympanostomy tubes + adenoidectomy vs. myringotomy + adenoidectomy	5 studies: 2 RCTs by person (N=130), 2 RCTs (by ears) (N=85), 1 NRCT (by ears) (N=146)	No difference in measured hearing between groups at 6 and 12 months and greater than 3 years.	Low for no difference
Adenoidectomy vs. no treatment or tympanostomy tubes	MA of 2 RCTs (N=153); MA of 3 RCTs (N=297)	Adenoidectomy better OME resolution at 6 months. The risk difference was 0.27 (95% CI, 0.13 to 0.42), measured through otoscopy and 0.22 (95% CI, 0.12 to 0.32) measured through tympanometry.	High for benefit
	MA of 3 RCTs (N=298)	Adenoidectomy better OME resolution at 12 months. The risk difference was 0.29 (95% CI, 0.19 to 0.39).	High for benefit
	1 study (N=237)	Adenoidectomy and myringotomy less mean time with effusion than myringotomy alone at 24 months: -0.76 standard mean difference (95% CI, -1.02 to -0.49)	Low for benefit
	1 study (N=237)	Adenoidectomy and myringotomy better hearing than with myringotomy alone at 24 months: -0.65 standard mean difference time with hearing level ≥ 20 worse ear (95% CI, -0.91 to -0.39)	Low for benefit
Oral steroids vs. controls	MA of 3 RCTs (N=106);	No difference in persisting OME at 1-2 months, (no antibiotics provided in either group) OR=0.55 (95% CI, 0.21 to 1.48) or	Low for no difference
Oral steroids + antibiotics vs. controls + antibiotics	MA of 3 RCTs (N=243)	No difference in persisting OME at 1-2 months (antibiotics provided to both groups) OR=0.75 (95% CI, 0.45 to 1.27)	Moderate for no difference
Topical intranasal steroids vs. controls	1 RCT (N=217)	No difference in OME cure rates at 1, 3, and 9 months	Low for no difference
	1 RCT (N=217)	No difference in hearing loss at 3 and 9 months	Low for no difference
Autoinflation vs. controls	MA of 2 RCTs (N=185)	Improvement in OME at ≤1month RR=3.84 (tympanometry change C2 to C1 or A) and RR=2.72 (tympanometry change B to C1 or A)	Low for benefit

Abbreviations: CI = confidence intervals; dB = decibels; NRCT = non-randomized controlled trial; MA = meta-analysis; N = number; OME; otitis media with effusion; OR = odds ratio; RCT = randomized controlled trial; vs. versus

Evidence concerning clinical outcome comparisons of tympanostomy tubes in relation to design (i.e., material, coatings, shape or size) or routes or techniques of insertion was considered insufficient because all single studies differed along one or more of these dimensions. Similarly, we found only one study examining different approaches to myringotomy without other interventions.

We found that tympanostomy tubes are more likely to decrease the time with persistent middle ear effusion over 1 year (high strength of evidence) and 2 years (moderate strength of evidence) compared to watchful waiting or myringotomy. Hearing, the more critical and patient-centered clinical outcome, was found to be superior with tympanostomy tubes as well, but for a shorter period of time, 4 to 6 months—measured by ears (high strength of evidence and 6 to 9 months—measured by child (moderate strength of evidence). Shorter time periods are likely to be more important for the youngest children (less than 3 years of age) who are still developing their speech and language skills but results were not available specifically for this age group. At increasingly longer outcomes, hearing differences between groups became smaller and not significantly different, 7 to 12 months—measured by ears (low strength of evidence for no difference), 12 and 18 months—measured by child (moderate strength of evidence for no difference). We found limited evidence concerning this treatment comparison in relation to improving OME recurrence, ear ventilation, or episodes of AOM and are unable to comment on these outcomes. We were also unable to disentangle comparisons between tympanostomy tubes and watchful waiting (separately from myringotomy). Because of the limited number of available studies, these two approaches were combined in the systematic review from which we obtained this meta-analysis data.[17]

We examined whether tympanostomy tubes or myringotomy are more likely to improve clinical outcomes when one or the other is added to adenoidectomy. We found no differences in hearing outcomes at any time points measured in five studies and concluded that because of this consistent finding, strength of evidence was low for no difference.

Adenoidectomy was found to be superior to no treatment or tympanostomy tubes in relation to improving the probability of OME resolution at 6 month and 1 year followup (strength of evidence high). Evidence was insufficient to determine if adenoidectomy also was superior to tympanostomy tubes in relation to hearing outcomes. In addition, adenoidectomy and myringotomy was found to be superior to myringotomy alone in relation to time with effusion and hearing outcomes at 24 months (strength of evidence low). Based on these findings, our review suggests that middle ear ventilation through tympanostomy tubes could be reconsidered when adenoids are removed. Given the similarity of hearing outcomes when tympanostomy tubes or myringotomy are added to adenoidectomy, our findings also suggest that it is unclear whether additional benefit is obtained from the myringotomy procedure.

In relation to nonsurgical interventions, we found evidence of oral steroids providing no short-term improvements in OME (at 1 to 2 months) with the addition of antibiotics (moderate for no difference) and without (low for no difference). One new low risk of bias study provided additional evidence that OME and hearing outcomes are not improved through the use of topical intranasal steroids at up to 9 months (low for no difference). These findings support the current American Academy of Pediatrics guidelines against the use of oral and intranasal steroids in treating OME in children.

Evidence concerning clinical outcomes related to autoinflation found improvement in relation to middle ear effusion at 1 month or less (low strength of evidence) but was insufficient for evaluating lengthier followup periods or in relation to hearing outcomes.

KQ 2: Health-Related Quality Of Life and Functional Outcomes

As in KQ 1, we sought evidence of the effectiveness of the various interventions to improve quality of life and subjective hearing, speech and language, or behavior.

Summary of Findings and Strength of Evidence

Of the evidence meeting our inclusion criteria for the review overall, only a small number of studies included data on health-related quality of life and functional outcomes. Evidence was limited to the following intervention comparisons: tympanostomy tubes versus watchful waiting or myringotomy and tympanostomy tubes plus adenoidectomy versus myringotomy plus adenoidectomy, and topical intranasal steroids versus control.

Table 39 summarizes the OME interventions on which we had low, moderate or high strength of evidence for clinical outcomes.

Table 39. Health-related quality of life and functional status

Intervention and Comparator	Number of Studies (Sample sizes)	Outcome and Results	Strength of Evidence
Tympanostomy tubes vs. watchful waiting or myringotomy	MA of 3 RCTs (N=394)	No difference in language comprehension at 6-9 months: Mean difference=0.09 (95% CI, -0.21 to 0.39)	Low for no difference
	MA of 3 RCTs (N=393)	No difference in language expression at 6-9 months: Mean difference=0.03 (95% CI, -0.41 to 0.49)	Low for no difference
Intranasal steroids vs. controls	1 study (N=144)	No difference in parent reported hearing difficulties at 3 and 9 months and median days with hearing loss at 3 months.	Low for no difference

Abbreviations: CI = confidence interval; MA = meta-analysis; N = number; vs. = versus

Language comprehension and language expression outcomes at 6-9 months were not significantly better among children with OME who received tympanostomy tubes than among those who were limited to watchful waiting or myringotomy (strength of evidence low). These findings correspond to clinical hearing outcomes also not being superior in the tympanostomy tube group after a similar period of followup. Evidence was insufficient to reach conclusions related to behavioral or quality of life outcomes differences for this treatment comparison.

Quality of life outcomes were measured in one small study comparing tympanostomy tubes and adenoidectomy versus myringotomy and adenoidectomy. Therefore, we considered the evidence to be insufficient to reach conclusions.

Parents' reports of their children's hearing difficulties were found in one low risk of bias study comparing intranasal steroids and controls (low SOE for no difference).

KQ 3: Harms Associated With Interventions to Treat Otitis Media With Effusion

We sought evidence of the potential harms or side effects that may occur with various treatment options. We considered such concerns as otorrhea, atrophy, tympanosclerosis, cholesteatoma and tissue granulation. Specifically, in relation to tympanostomy tubes we considered otorrhea and perforation and such problems as diarrhea and nasal stinging in relation to steroid treatment. Table 40 summarizes the OME interventions on which we had low, moderate or high strength of evidence for harms outcomes.

Table 40. Strength of evidence for harms or tolerability of interventions

Intervention and Comparator	Number of Studies (Sample sizes)	Outcome and Results	Strength of Evidence
Tympanostomy tubes vs. watchful waiting or myringotomy	5 studies (N=1129)	Tympanosclerosis occurred more frequently in ears that had tympanostomy tubes, based on examinations after the tubes had been extruded.	Low for harms
	4 studies (N=960)	Otorrhea occurred more frequently in ears with tympanostomy tubes.	Low for harms
Tympanostomy tubes plus adenoidectomy vs adenoidectomy alone or with myringotomy	2 studies (N=237)	Tympanosclerosis occurred more frequently in ears with tympanostomy tubes.	Low for harms
Topical nasal steroids vs. control	1 study (N=170)	No difference in mild adverse events such as nasal stinging, dry throat, and cough	Low for no difference

Abbreviations: N = number; NR = not reported; SR = systematic review; vs. = versus

We found consistent evidence that tympanosclerosis was more common in children who had tympanostomy tubes than those who were watched or who had myringotomy, both in comparison to adenoidectomy and watchful waiting or myringotomy (strength of evidence low). Otorrhea was also more common among ears with tympanostomy tubes (strength of evidence low).

The systematic review concerning nasal steroids[56] we included as evidence found few mild adverse events in the studies they reviewed. Similarly, one new study[57] found no differences between groups in relation to stinging nose, nose bleed, dry throat, or cough. We therefore conclude that mild adverse events are not significantly higher through the use of topical nasal steroids (low for no difference). However, evidence was insufficient to reach conclusions related to mild adverse events concerning oral steroids and serious adverse events from oral or topical steroids.

We found no evidence concerning harms from adenoidectomy, including no data on any risks from having a surgical procedure. Similarly, we found insufficient evidence concerning the surgical risks from the insertion of tympanostomy tubes or myringotomy procedures.

KQ 4: Outcomes for Important Patient Subgroups

We aspired to differentiate treatment efficacy or harms for key subgroups characterized by clinical or sociodemographic factors (such as age). For example, clinicians often treat children with preexisting hearing deficiencies, Down syndrome or cleft palate differently than they would manage children who are do not have such coexisting or congenital conditions and are otherwise following a typical development trajectory. Despite the important clinical and social questions that arise for children or adults in such subgroups, we could not identify studies that included most of our subgroups of interest.

Two studies examined different subgroups—children with sleep apnea and adults with OME (Table 41) Vlastos et al. performed a study specifically on children with sleep apnea and OME. Among children with sleep apnea, all of whom had adenoidectomy to treat that condition, we found insufficient evidence to reach conclusions in terms of any measured outcomes.[52] A study of autoinflation that was included in the systematic review[19] included as evidence found differences in rates of recovery between those receiving autoinflation and those who were in the control group (strength of evidence low).

Table 41. Strength of evidence for subgroups

Intervention and Comparator	Number of Studies (Sample sizes)	Subgroup and Results	Strength of Evidence
Tympanostomy tubes + adenoidectomy vs. myringotomy + adenoidectomy	1 RCT (N=52)	Children with sleep apnea and OME: no difference between tympanostomy tubes and myringotomy outcomes (all children had received adenoidectomies)	Insufficient (one study)
Autoinflation vs. control	1 RCT (N=198)	Adults (16-75) with OME: differences between groups in composite measure of recovery (otoscopy, tympanometry, audiometry) at end of tx and 50 days after tx, 44% to 47%.	Low for benefit

Abbreviations: TT = tympanostomy tubes; OME = otitis media with effusion; tx = treatment

KQ 5: Health Care Factors

We found no studies that examined issues related to health insurance coverage, physician specialty, type of facility of the provider, geographic location of patients, presence or absence of continuity of care, or prior use of pneumococcal virus inoculation. Evidence is thus insufficient for all such considerations.

Findings in Relation to What Is Already Known

The preponderance of the evidence included in this systematic review was obtained from recently completed reviews. Four of these were conducted by the Cochrane Collaboration[18-20] and the fifth was sponsored by the Swedish government.[15] We sought to determine if the inclusion of non-RCT evidence (excluded from the Cochrane reviews) and newer trials would affect their findings. We also sought to obtain answers to questions not addressed in these reviews, such as the comparative effectiveness of different types of tympanostomy tubes or different approaches to myringotomy as well as the use of CAM therapies in treating OME and the value of watchful waiting. Lastly, we sought evidence concerning populations not addressed in these reviews, such as findings specific to very young children, adults (an adult nominated the review) and children at greater risk for hearing deficiencies or developmental delays due to preexisting comorbidities.

Overall, we found little new evidence that was not included in the earlier reviews. We found one new low risk of bias RCT concerning topical steroid treatment.[56, 57] However, new evidence from non-randomized trials and observational studies did not add appreciably to our understanding of these treatment comparisons. Nor were we able to uncover much evidence regarding special populations. Thus our conclusions are largely a compilation of those that have been made in the previous systematic reviews.

Implications for Clinical and Policy Decisionmaking

The evidence from this review largely reiterates that provided in past systematic reviews of treatments for OME. We did not find new evidence to refute the findings of Thomas et al.[20] concerning the lack of effectiveness of oral and intranasal steroids as treatment for OME and have provided additional evidence from a recent RCT of Williamson et al.[56, 57] that intranasal steroids were not found to be effective. This evidence supports the recommendations made by the American Academy of Pediatrics that neither oral nor topical nasal steroids be used in the treatment repertoire for OME.

We found some evidence that tympanostomy tubes are effective for reducing effusion and for improving hearing but their effect is limited, no doubt a consequence of the fact that effusion

resolves even if untreated. However, we do not have evidence regarding which types of tubes or tube routines for insertion are more beneficial for reducing fluid and mitigating harms and whether there are differences for younger versus older individuals. We still do not know for what age child it is most deleterious for fluid to remain untreated. Nor do we know whether sub-populations of children with cleft palate or Down syndrome need to follow a different treatment course as typically developing children.

Overall, there was evidence that children with tympanostomy tube placement for OME are more likely to have resolution of middle ear effusion for up to 2 years after the procedure. There was a similar difference noted for hearing loss up to 6 months after tube placement. This difference and the physiological and developmental plausibility that the hearing loss could worsen speech and language outcomes short and long term has driven clinicians to intervene on prolonged OME. Because in the longer term, effusions resolve in the vast majority of patients without any intervention, a key clinical decision concerns the length of time that mild to moderate hearing loss needs to be present to negatively impact important outcomes, and how these may differ for individuals at different developmental stages and ages.

Many primary care providers refer patients with prolonged effusion and mild to moderate hearing loss to otolaryngologists for placement of tympanostomy tubes, after 3 months of effusion. However, our synthesis of the available evidence, did not find functional or quality of life differences between subjects who had tympanostomy tube placement versus watchful waiting for OME with only mild to moderate hearing loss.

Currently, many children with craniofacial syndromes or underlying hearing loss have tympanostomy tubes placed either prophylactically (e.g. cleft palate) or at a very low threshold of time that effusion is present. We found no evidence specific to these populations to either support or refute those practices.

Adenoidectomy alone is a more effective treatment for middle ear effusion relative to watchful waiting or tympanostomy tube placement. However, surgery for adenoidectomy is more invasive and raises concern that it may impose more short term complications than tympanostomy tubes but we did not find evidence describing or quantifying the risk.

For clinical questions that have insufficient evidence to provide confident answers, clinicians will need to continue to rely on the expert opinions contained within clinical practice guidelines, clinician experience, and individual patient and family level shared decisionmaking.

Applicability

As noted, during the review process we systematically abstracted key factors that may affect the applicability of the evidence base. We identified these key factors a priori, defining applicability as “the extent to which the effects observed in published studies are likely to reflect the expected results when a specific intervention is applied to the population of interest under real-world conditions.”[60]

Population

Findings about all interventions are likely to be applicable to non-infant, otherwise healthy children. However, the evidence base is limited for adults and for infant children or children with major coexisting or congenital conditions who may be at risk of having OME for longer periods of time, such as those with cleft palate or Down syndrome, and those who may be more sensitive to hearing loss such as those with preexisting hearing loss. Despite our goal of examining outcomes in young children, adults, and individuals with co-morbidities, we were unable to find

sufficient, if any studies on these populations, and so do not know if the relative efficacy of treatment comparisons will be similar for them.

Intervention and Comparators

We present evidence on all of the commonly used treatments for OME, including tympanostomy tubes, myringotomy, adenoidectomy, and watchful waiting. We present evidence on oral and intranasal steroids because, while not currently recommended in major guidelines or approved by the Food and Drug Administration for treating children with OME, our TEP believed that it was still a commonly used intervention. We also include autoinflation, because, while not typically used in the U.S., offers an alternative non-invasive treatment strategy for older children and adults. We sought to determine the applicability of "watchful waiting" but because we were unable to isolate this approach from myringotomy, we were unable to determine the population and length of time that this approach may be most useful.

Outcomes

We did not limit the outcomes of interest but rather took a broad view of what kinds of benefits might occur with the treatments. We targeted clinical health outcomes, health care utilization as well as functional and quality of life outcomes. However, the bulk of the literature only examined whether the interventions reduced OME or improved hearing. A few studies included in the systematic reviews examined language development and behavior problems, and a few of the systematic reviews and two recent RCTs examined quality of life outcomes. No studies focused on parental and patient satisfaction with care or heath care utilization. Thus, there is little that we can say about these other important outcomes. We acknowledge the central role that continued effusion and hearing play in these other outcomes, yet the broad range of outcomes are important in their own right. Moreover, investigators chose different measures to index many of these outcomes (e.g., quality of life), so that even when we had two or more studies reporting an outcome, we could not perform quantitative summaries. Our lack of evidence in these areas parallels the conclusions reported in the previous AHRQ systematic review[2] of late effects of OME that failure to reach conclusions about effects of OME on long term speech and language are in part due to lack of uniformity in instrumentation. The most robust test of whether a particular treatment was efficacious is to examine treatment comparison pre-post differences; however, this was not uniformly done. There was a similar lack of uniformity in reporting harms and failure to provide information about patients in all arms of a study.

Time Frames

Studies varied in relation to their length of followup periods. Many included studies measured results between 3 months and 12 months following treatment. However, for some comparisons, such as differences between types of tympanostomy tubes and autoinflation and controls, followup was generally shorter. However, for any given comparison, time frames were rarely uniform, making cross-study integration difficult.

Settings

Studies were conducted in clinical setting and generally included populations from the United States and European countries. There were a few studies conducted in developing countries such as Bulgaria and non-Western countries, including Japan, Egypt, and Iran.

Limitations of the Review Process

As noted previously, we limited the synthesis to low and medium risk of bias. Given the limitations of the included studies and their applicability to other contexts, however, including high risk-of-bias studies would likely have increased the pool of evidence without resulting in more actionable evidence.

Limitations of the Evidence Base

The evidence base was limited by the decision to restrict studies to those that examined the treatments in individuals with OME. Much of the literature only indicated that the study participants had otitis media without providing information about the type or included a mixed sample of individuals with AOM and OME. When investigators analyzed the OME samples separately, we included the study, but this was the exception rather than the rule. Thus, the body of research we included was restricted because of the lack of specificity in populations covered.

The evidence base was further limited by a lack of low risk of bias studies. Overall, we rated only one of the 20 new studies included in our analysis as low risk of bias. Some of the major reasons we rated studies as medium risk of bias included lack of information regarding randomization, blindedness of providers, and high rates of attrition. Overall, the better studies were in the systematic reviews that we included in our review.

The evidence regarding management of OME is further confounded by a variety of methodological limitations. Studies employ a range of criteria for diagnostic inclusion as well as a wide variety in outcomes measures and scant description of how those measures are obtained. Even when measures are similar there are often variations in when they are collected, the qualifications of those conducting the measures, and the ways in which they are reported. Most studies include a wide range of children (2-14 years of age). It is likely that those children age 6 years or more who experience OME are at different risk than those who experience episodes of OME in their early years but who experience no episodes after age 4. Including all children in studies may mask the benefits of treatment of individuals at varying ages. Because of this heterogeneity, we were unable to perform any quantitative analysis.

Few studies provided a power analysis so making it difficult to interpret failure to find differences. However, we suspect that power was low given the relatively small, heterogeneous samples and extensive attrition in some trials.

Research Gaps

Given the severe limitations of the evidence base—both gaps in study topics (interventions, appropriate outcomes, relevant populations) and in methods, we have several recommendations for future directions.

Gaps in Subgroups Studied

Additional research needs to determine the appropriate criteria and waiting period prior to surgical intervention with children Although children 3 years or older may be able to tolerate a mild-moderate hearing loss for a period of 3-6 months or longer without risk to language outcomes, the effect of the same hearing loss on children 24 months or younger is unknown. Children who are in the earliest stages of phonological and language development are likely to be much more vulnerable to the effects of OME induced hearing loss than children 3 years of age or older. A hearing loss of any degree creates a barrier to full access to the auditory signal. Infants and toddlers who are learning the rules that govern language comprehension and production are likely to be vulnerable to the hearing loss imposed by OME. Research in the area of infant speech perception and later outcomes has demonstrated that babies who were able to distinguish between two simple vowels /i/ (tea) and /u/ (two) at 6 months of age had larger vocabularies when they were 18 an 24 months old than those who could not. Early vocabulary development is important because it is one of the strongest predictors of academic achievement.[61]

In many instances children younger than 2 years of age are underrepresented in studies and when they are included the results are not portioned by age. We recommend that RCT's that include children at the most vulnerable ages examine effects of OME on morphosyntactical development (0-36 months) and report results partitioned by age groups reflecting developmental vulnerability.

Evidence regarding the impact of interventions for OME in at-risk sub-populations is virtually non-existent. Children with a variety of developmental or sensory delays are usually excluded from studies investigating treatment outcomes for OME, including those with Down syndrome, permanent sensorineural or conductive hearing loss, craniofacial anomalies affecting Eustachian tube function such as cleft palate, ciliary dyskinesia, and children on the autism disorder spectrum. All of these subgroups are at-risk for development of speech and language because of these comorbidities and the addition of a 15-20 dB hearing loss due to OME increases their vulnerability. Although RCTs may not be feasible because of ethical concerns, as well as the relatively low incidence of these conditions, carefully controlled observational studies are needed to guide management of OME in these subgroups.

Despite the high prevalence of OME in children with cleft palate,[62, 63] we found no evidence on treatment of OME in this population that met our inclusion criteria. Studies were mainly excluded because tympanostomy tube placement occurred prophylactically during other craniofacial surgery and was not limited to just to children with diagnoses of OME. A recent 2009 systematic review, conducted by Ponduri and colleagues,[64] assessed evidence on OME related symptoms and hearing, speech and language outcomes in children with cleft palate who received early placement of tympanostomy tubes. Only three of the studies included in the systematic report are limited to children who were diagnosed with OME at the time of initial assessment.[65-67] These studies were excluded for wrong population (included children with suppurative otitis media), wrong study design (case series-no comparator), and wrong publication type (the study was not available in English), respectively.

We identified one ongoing study with adult participants. The Children's Hospital of Pittsburgh in collaboration with The National Institute on Deafness and Other Communication Disorders (NIDCD) and University of Pittsburgh, are currently conducting an observational study of adults who have received tympanostomy tubes for the treatment of chronic OME and/or Eustachian tube function.[68] The described aims of the study are to explore the relationship

between the results of standard Eustachian tube function tests, and those for the anatomy and function of the Eustachian tube of adults with a disease condition that is likely to be caused by poor Eustachian tube function. The investigators plan to evaluate study participants for a number of disease conditions (examples: nasal allergy, acid reflux disease) that are related to poor Eustachian tube function. The investigators hypothesize that if the results of the standard Eustachian tube function tests can be explained by the physical anatomy of Eustachian tube and back of the nose, the functional anatomy (e.g. movements during swallowing) of the Eustachian tube, or the presence of allergy and/or acid reflux disease, these findings will encourage development and/or use of new medical and/or surgical treatments to improve Eustachian tube function and outcomes associated with middle-ear diseases.

Gaps in Outcomes Measured (Benefits or Harms)

As indicated previously, outcomes were mainly limited to resolution of OME and hearing. As important as these are because they can be easily measured, we do not know to what extent they are correlated with functional outcomes such as speech and language development and quality of life. Also, we found no information on how treatment choice during one OME episode affected later utilization of health care services. We believe that one area for future research is to establish whether treatments can impact these outcomes.

For instance, we had targeted auditory processing as an outcome of interest because research has demonstrated that OME can affect skills such as binaural auditory perception,[69] and speech recognition in noise.[61, 70] Presumably, these skills affect children's ability to attend to instruction in noisy classrooms. One small study by Hall et al.[71] found that tympanostomy tubes can improve one measure of auditory processing, but the recovery period is protracted. Hearing is necessary for auditory processing but even when hearing returns to normal, auditory perception can still be impaired. This study was not a trial and only included a small number of children. However, replication with larger samples as part of a clinical trial would be extremely useful as there is relevance for daily life.

There were no studies that examined either health care utilization or parent satisfaction with care. Whether any of these treatments under question reduce time spent at the physician's office, along with costs associated with loss of productivity are not known. Anecdotally, we know that parents often request tympanostomy tubes because they hope that it will reduce time that their children are ill. The unexamined issue is whether receiving tubes or another treatment options affects these secondary outcomes.

We had few conclusions with regard to harms, in part because the evidence base was so limited. Future studies should aim to examine a uniform body of harms for all patients. Some of the treatment complications are rare (e.g., cholesteatoma) or complications from adenoidectomy surgery.

Gaps in Interventions

This review provided little evidence regarding different types of tubes or routines for insertion. However, an ongoing Swedish trial plans to enroll 400 children between the ages of 1 and 10 in an RCT comparing complications from four types of tubes, two different materials, and two different shapes.[72] The comparisons described are the Shepard tube (double flanged, Fluoroplastic) versus Donaldson tube (double flanged, silicone); Straight tube (single flanged, Fluoroplastic) versus Armstrong (single flanged, silicone); Armstrong (single flanged, silicone) versus Donaldson tube (double flanged, silicone); and Straight tube (single flanged,

Fluoroplastic) versus Shepard tube (double flanged, Fluoroplastic). Outcomes include time to complete expulsion of the tube from the tympanic membrane, and various harms including persistent tympanic membrane perforation, need for tube extraction (i.e., pain or infection), pain leading to health care contact, tube related ear infection, obstruction of tube and presence of myringosclerosis. The study will include both children with recurrent acute otitis media (RAOM) and OME, but presentation of the results (i.e., complications in OME and RAOM reported separately) could inform best practices in tympanostomy tube choice for children with OME.

Despite increasing interest by the public in alternatives to surgical interventions or traditional medical management, an exhaustive review of the literature failed to identify any RCTs regarding CAM treatments. The need for carefully conducted investigations of CAMs, including dietary modifications, has been identified. We identified an ongoing, and potentially promising RCT that addresses the benefit of dietary modification in the treatment of OME.[73] The study, being conducted at the Department of Otolaryngology at the University Hospital at the University of Missouri-Columbia, hopes to provide evidence that standard treatment options for chronic OME in children should involve food allergy assessment and, when indicated, subsequent dietary modifications in addition to standard surgical procedures. Additionally, investigators are seeking evidence to assess whether adenoidectomy is of added benefit in a treatment course of surgical intervention and dietary modification. Investigators plan to measure recurrence of OME in two treatment groups, bilateral myringotomy with tympanostomy tubes, in conjunction with food allergy testing and management; and bilateral myringotomy with tympanostomy tubes in conjunction with adenoidectomy, and food allergy testing and management. The incidence of recurrent OME episodes in all trial groups will be recorded at 3 month intervals until tube expulsion, with a further year of followup evaluations at 3 month intervals. At the time of this report, this study is listed as recruiting.[73]

Several studies have found high pepsin or pepsinogen, a component of stomach fluid, in the middle ear fluid of children with chronic middle ear fluid and some researchers believe gastroesophageal reflux disease (GERD) may be a cause of OME.[74] We identified two unpublished trials which evaluate the treatment of OME in children with Proton Pump Inhibitors, anti-reflux medication. One study of chronic OME is listed as completed,[72] with December 2009 reported as the final data collection date for the primary outcome measure(s), however we were unable to identify any related publications, and to the best of our ability do not know of any publications on the outcomes of this study. An ongoing pilot study of anti-acid treatment for children and adolescents with OME lists the completion date for data collection of primary outcome measures as April, 2012. While the primary goal of this study is to collect data for the purposes of calculating sample size and recruitment rates required for a larger clinical trial, the secondary outcomes could potentially be of particular interest to the field; these outcomes include degree of hearing improvement, complications of OME (e.g. recurrent OME, surgery) and side effects of lansoprazole. The larger clinical trial that is set to follow could provide a clearer picture of the role gastric reflux in OME, and inform treatment decisions although recent evidence of risks associated with proton-pump inhibitor use in children will need to be incorporated into treatment decisions.[75]

A prospective cohort study, ongoing since 2006, in children 3 to 6 years of age who underwent tympanostomy tube insertion for chronic OME aims to determine if Eustachian tube function tests and gas exchange tests can be used to successfully predict whether a child who has tympanostomy tubes surgically placed in their eardrum because of middle-ear disease will redevelop the disease again after the tube is nonfunctional and/or expelled.[76] The investigators

state that the results of this study will be used to support or contest components of existing models of middle ear pressure-regulation, and to develop test protocols for risk assignments of disease recurrence in individual ears after tympanostomy tubes become nonfunctioning or extruded.

Deficiencies in Methods

Meta-analyses can be used to strengthen the power for finding effects in studies where there are a limited number of participants in individual studies. However, in many cases meta-analysis is hampered by methodological differences in studies. In particular, vastly different outcome measures and outcome assessment times limits the studies that can be included. If there was agreement among investigators about outcomes to be included in studies, more quantitative analyses could be performed. At a minimum, uniform time points for outcome assessments and consistency in measures of hearing would make the task of combining research easier.

Conclusions

Overall, we found uneven bodies of evidence across treatment comparisons. In particular, limited data was available on primary patient-centered outcomes: hearing loss, functional limitations, and quality of life. Additional research is needed to develop a sufficient evidence base to support treatment decisions, particular in subpopulations.

References

1. Stool SE, Berg AO, Berman S, et al. Otitis Media With Effusion in Young Children. Agency for Health Care Policy and Research Clinical Practice Guideline, No. 12. Rockville, MD: 1994. AHCPR Publication No. 94-0622.
2. Shekelle P, Takata G, Chan LS, et al. Diagnosis, History, and Late Effects of Otitis Media With Effusion. Agency for Healthcare Research and Quality Evidence Report/Technology Assessment No. 55 (Prepared by Southern California Evidence-based Practice Center under Contract No 290-97-0001, Task Order No. 4). Rockville, MD: 2003. AHRQ Publication No. 03-E023.
3. Cantekin EI, Bluestone CD. Membrane ventilating tube for the middle ear. Ann Otol Rhinol Laryngol. 1976 Mar-Apr; 85(2 Suppl 25 Pt 2):270-6. PMID: 1267358.
4. Klein JO, Pelton S. Otitis media with effusion (serous otitis media) in children. In: Basow DS, ed UpToDate. Waltham, MA; 2012.
5. Paradise JL, Rockette HE, Colborn DK, et al. Otitis media in 2253 Pittsburgh area infants: prevalence and risk factors during the first two years of life. Pediatrics. 1997 Mar; 99(3):318-33. PMID: 9041282.
6. Tos M. Epidemiology and natural history of secretory otitis. Am J Otol. 1984 Oct; 5(6):459-62. PMID: 6542752.
7. Williamson IG, Dunleavey J, Bain J, et al. The natural history of otitis media with effusion--a three-year study of the incidence and prevalence of abnormal tympanograms in four South West Hampshire infant and first schools. J Laryngol Otol. 1994 Nov; 108(11):930-4.
8. National Collaborating Centre for Women’s and Children’s Health. Surgical Management of Otitis Media With Effusion in Children. NICE Clinical Guideline 60. London: National Institute for Health and Clinical Excellence; February 2008.
9. Maw AR. Otitis media with effusion (glue ear). In: Kerr AG, Groves J, Evans NG, eds. Scott-Browan's Otolaryngology. 5th ed. London: Butterworths; 1987:159-76.
10. Finkelstein Y, Ophir D, Talmi YP, et al. Adult-onset otitis media with effusion. Arch Otolaryngol Head Neck Surg. 1994 May; 120(5):517-27. PMID: 8172703.
11. Gravel JS, Wallace IF. Effects of otitis media with effusion on hearing in the first 3 years of life. J Speech Lang Hear Res. 2000 Jun; 43(3):631-44. PMID: 10877434.
12. Rosenfeld RM, Culpepper L, Doyle KJ, et al. Clinical practice guideline: Otitis media with effusion. Otolaryngol Head Neck Surg. 2004 May; 130(5 Suppl):S95-118. PMID: 15138413.
13. Sano S, Kamide Y, Schachern PA, et al. Micropathologic changes of pars tensa in children with otitis media with effusion. Arch Otolaryngol Head Neck Surg. 1994 Aug; 120(8):815-19. PMID: 8049041.
14. Madell JR. Impact of Otitis Media on Auditory Function. In: Rosenfeld RM, Bluestone CD, eds. Evidence-Based Otitis Media. Canada: B.C. Decker Inc.; 1999:337-51.
15. Hellstrom S, Groth A, Jorgensen F, et al. Ventilation tube treatment: a systematic review of the literature. Otolaryngol Head Neck Surg. 2011 Sep; 145(3):383-95. PMID: 21632976.
16. Williamson I. Otitis media with effusion in children. Clinical Evidence. 2011; pii: 0502PMID: 21477396.
17. Browning GG, Rovers MM, Williamson I, et al. Grommets (ventilation tubes) for hearing loss associated with otitis media with effusion in children. Cochrane Database Syst Rev. 2010; (10):CD001801. PMID: 20927726.
18. van den Aardweg MT, Schilder AG, Herkert E, et al. Adenoidectomy for otitis media in children. Cochrane Database Syst Rev. 2010; (1):CD007810. PMID: 20091650.
19. Perera R, Haynes J, Glasziou P, et al. Autoinflation for hearing loss associated with otitis media with effusion. Cochrane Database Syst Rev. 2006; (4):CD006285. PMID: 17054290.

20. Thomas CL, Simpson S, Butler CC, et al. Oral or topical nasal steroids for hearing loss associated with otitis media with effusion in children. Cochrane Database Syst Rev. 2006; (3):CD001935. PMID: 16855980.

21. Griffin GH, Flynn C, Bailey RE, et al. Antihistamines and/or decongestants for otitis media with effusion (OME) in children. Cochrane Database Syst Rev. 2006; (4):CD003423. PMID: 17054169.

22. Deutsch ES, Cook SP, Shaha S, et al. Duration of patency of laser-assisted tympanic membrane fenestration. Arch Otolaryngol Head Neck Surg. 2003 Aug; 129(8):825-8. PMID: 12925339.

23. . Which antibiotic for acute otitis media? Drug Ther Bull. 1974 Jul 5; 12(14):53-4. PMID: 4142433.

24. Hassmann E, Skotnicka B, Baczek M, et al. Laser myringotomy in otitis media with effusion: long-term follow-up. Eur Arch Otorhinolaryngol. 2004 Jul; 261(6):316-20. PMID: 14551787.

25. Lin SH, Lai CC, Shiao AS. CO2 laser myringotomy in children with otitis media with effusion. J Laryngol Otol. 2006; 120(3):188-92. PMID: 2009230255. Language: English. Entry Date: 20080201. Revision Date: 20091218. Publication Type: journal article.

26. Sedlmaier B, Jivanjee A, Gutzler R, et al. Ventilation time of the middle ear in otitis media with effusion (OME) after CO2 laser myringotomy. Laryngoscope. 2002 Apr; 112(4):661-8. PMID: 12150520.

27. Rosenfeld RM, Bluestone CD, eds. Evidence-Based Otitis Media. Ontario: BC Decker Inc; 1999.

28. Wahl RA, Aldous MB, Worden KA, et al. Echinacea purpurea and osteopathic manipulative treatment in children with recurrent otitis media: a randomized controlled trial. BMC Complement Altern Med. 2008; 8:56. PMID: 18831749.

29. van Zon A, Schilder AGM, van der Heijden GJ. Antibiotics for otitis media with effusion in children (Protocol). Cochrane Database of Systematic Reviews 2011.

30. Owens DK, Lohr KN, Atkins D, et al. AHRQ series paper 5: grading the strength of a body of evidence when comparing medical interventions--agency for healthcare research and quality and the effective health-care program. J Clin Epidemiol. 2010 May; 63(5):513-23. PMID: 19595577.

31. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol. 2009 Oct; 62(10):1006-12. PMID: 19631508.

32. Lerman SJ, Kucera JC, Brunken JM. Nasopharyngeal carriage of antibiotic-resistant Haemophilus influenzae in healthy children. Pediatrics. 1979 Sep; 64(3):287-91. PMID: 314623.

33. Hartling L, Bond K, Harvey K, et al. Developing and testing a tool for the classification of study designs in systematic reviews of interventions and exposures. Methods Research Report Rockville, MD: Agency for Healthcare Research and Quality; December 2010. AHRQ Publication No. 11-EHC-007.

34. Owens DK, Lohr KN, Atkins D, et al. AHRQ series paper 5: grading the strength of a body of evidence when comparing medical interventions--Agency for Healthcare Research and Quality and the Effective Health-Care Program. J Clin Epidemiol. 2010 May; 63(5):513-23. PMID: 19595577.

35. Ovesen T, Felding JU, Tommerup B, et al. Effect of N-acetylcysteine on the incidence of recurrence of otitis media with effusion and re-insertion of ventilation tubes. Acta Otolaryngol Suppl. 2000; 543:79-81. PMID: 10908985.

36. Wielinga EW, Smyth GD. Comparison of the Goode T-tube with the Armstrong tube in children with chronic otitis media with effusion. J Laryngol Otol. 1990 Aug; 104(8):608-10. PMID: 2230552.

37. Iwaki E, Saito T, Tsuda G, et al. Timing for removal of tympanic ventilation tube in children. Auris Nasus Larynx. 1998 Dec; 25(4):361-8. PMID: 9853658.

38. Slack RW, Gardner JM, Chatfield C. Otorrhoea in children with middle ear ventilation tubes: a comparison of different types of tubes. Clin Otolaryngol Allied Sci. 1987 Oct; 12(5):357-60. PMID: 3427799.

39. Licameli G, Johnston P, Luz J, et al. Phosphorylcholine-coated antibiotic tympanostomy tubes: are post-tube placement complications reduced? Int J Pediatr Otorhinolaryngol. 2008 Sep; 72(9):1323-8. PMID: 18635268.

40. Abdullah VA, Pringle MB, Shah NS. Use of the trimmed Shah permavent tube in the management of glue ear. J Laryngol Otol. 1994 Apr; 108(4):303-6. PMID: 8182314.

41. McRae D, Gatland DJ, Youngs R, et al. Aspiration of middle ear effusions prior to grommet insertion an etiological factor in tympanosclerosis. J Otolaryngol. 1989 Aug; 18(5):229-31. PMID: 2769837.

42. Hampal S, Flood LM, Kumar BU. The mini-grommet and tympanosclerosis. J Laryngol Otol. 1991 Mar; 105(3):161-4. PMID: 2019798.

43. Heaton JM, Bingham BJ, Osborne J. A comparison of performance of Shepard and Sheehy collar button ventilation tubes. J Laryngol Otol. 1991 Nov; 105(11):896-8. PMID: 1761941.

44. Mandel EM, Rockette HE, Bluestone CD, et al. Myringotomy with and without tympanostomy tubes for chronic otitis media with effusion. Arch Otolaryngol Head Neck Surg. 1989 Oct; 115(10):1217-24. PMID: 2789777.

45. Koopman JP, Reuchlin AG, Kummer EE, et al. Laser myringotomy versus ventilation tubes in children with otitis media with effusion: a randomized trial. Laryngoscope. 2004 May; 114(5):844-9. PMID: 15126741.

46. Brown MJ, Richards SH, Ambegaokar AG. Grommets and glue ear: a five-year follow up of a controlled trial. J R Soc Med. 1978 May; 71(5):353-6. PMID: 650647.

47. Austin DF. Adenotonsillectomy in the treatment of secretory otitis media. Ear Nose Throat J. 1994 Jun; 73(6):367-9, 73-4. PMID: 8076535.

48. Lildholdt T. Unilateral grommet insertion and adenoidectomy in bilateral secretory otitis media: preliminary report of the results in 91 children. Clin Otolaryngol Allied Sci. 1979 Apr; 4(2):87-93. PMID: 455774.

49. D'Eredità R, Shah UK. Contact diode laser myringotomy for medium-duration middle ear ventilation in children. Int J Pediatr Otorhinolaryngol; 2006. p. 1077-80.

50. Popova D, Varbanova S, Popov TM. Comparison between myringotomy and tympanostomy tubes in combination with adenoidectomy in 3-7-year-old children with otitis media with effusion. Int J Pediatr Otorhinolaryngol. 2010 Jul; 74(7):777-80. PMID: 20399511.

51. Shishegar M, Hoghoghi H. Comparison of adenoidectomy and myringotomy with and without tube placement in the short term hearing status of children with otitis media with effusion: a preliminary report. Iranian Journal of Medical Sciences; 2007. p. 169-72.

52. Vlastos IM, Houlakis M, Kandiloros D, et al. Adenoidectomy plus tympanostomy tube insertion versus adenoidectomy plus myringotomy in children with obstructive sleep apnoea syndrome. J Laryngol Otol. 2011 Mar; 125(3):274-8. PMID: 21205368.

53. Tos M, Stangerup SE. Hearing loss in tympanosclerosis caused by grommets. Arch Otolaryngol Head Neck Surg. 1989 Aug; 115(8):931-5. PMID: 2751852.

54. Ragab SM. The effect of radiofrequency and mitomycin C on the closure rate of human tympanostomy. Otol Neurotol. 2005 May; 26(3):355-60. PMID: 15891633.

55. Szeremeta W, Parameswaran MS, Isaacson G. Adenoidectomy with laser or incisional myringotomy for otitis media with effusion. Laryngoscope. 2000 Mar; 110(3 Pt 1):342-5. PMID: 10718416.

56. Williamson I, Benge S, Barton S, et al. Topical intranasal corticosteroids in 4-11 year old children with persistent bilateral otitis media with effusion in primary care: double blind randomised placebo controlled trial. BMJ. 2009; 339:b4984. PMID: 20015903.

57. Williamson I, Benge S, Barton S, et al. A double-blind randomised placebo-controlled trial of topical intranasal corticosteroids in 4- to 11-year-old children with persistent bilateral otitis media with effusion in primary care. Health Technol Assess. 2009 Aug; 13(37):1-144. PMID: 19671372.

58. Dingle AF, Flood LM, Kumar BU, et al. The mini-grommet and tympanosclerosis: results at two years. J Laryngol Otol. 1993 Feb; 107(2):108-10. PMID: 8496640.

59. Lesinskas E. Factors affecting the results of nonsurgical treatment of secretory otitis media in adults. Auris Nasus Larynx. 2003 Feb; 30(1):7-14. PMID: 12589843.

60. Atkins D, Chang SM, Gartlehner G, et al. Assessing applicability when comparing medical interventions: AHRQ and the Effective Health Care Program. J Clin Epidemiol. 2011 Nov; 64(11):1198-207. PMID: 21463926.

61. Kuhl PK, Coffey-Corina S, Padden DM, et al. Early speech perception and later language development: Implications for the "Critical Period". Language Learning and Development. 2005; 1(3-4):237-64.

62. Grant HR, Quiney RE, Mercer DM, et al. Cleft palate and glue ear. Arch Dis Child. 1988 Feb; 63(2):176-9. PMID: 3348665.

63. Paradise JL, Bluestone CD, Felder H. The universality of otitis media in 50 infants with cleft palate. Pediatrics. 1969 Jul; 44(1):35-42. PMID: 5795401.

64. Ponduri S, Bradley R, Ellis PE, et al. The management of otitis media with early routine insertion of grommets in children with cleft palate -- a systematic review. Cleft Palate Craniofac J. 2009 Jan; 46(1):30-8. PMID: 19115800.

65. Paradise JL, Bluestone CD. Early treatment of the universal otitis media of infants with cleft palate. Pediatrics. 1974 Jan; 53(1):48-54. PMID: 4809194.

66. Moller P. Selective use of ventilating tubes in the treatment of secretory otitis media and retractions of the ear drum. Acta Otolaryngol (Stockh). 1982; 93(s386):158-60

67. Zheng Q, Xu H, He Y. [Effects of tympanotomy and pressure equilibrium tube insertion during palatoplasty on prognoses of otitis media with effusion]. Hua Xi Kou Qiang Yi Xue Za Zhi. 2003 Feb; 21(1):28-30. PMID: 12674616.

68. ClinicalTrials.gov [Internet]. Identifier: NCT01251432 Risk Factors and Potential Causes of Eustachian Tube Dysfunction in Adults. Bethesda (MD): National Library of Medicine (US); 2010. Available at: http://clinicaltrials.gov/ct2/show/NCT01251432. Accessed 2012 May 22.

69. Hall JW, 3rd, Grose JH, Pillsbury HC. Long-term effects of chronic otitis media on binaural hearing in children. Arch Otolaryngol Head Neck Surg. 1995 Aug; 121(8):847-52. PMID: 7619408.

70. Schilder AG, Snik AF, Straatman H, et al. The effect of otitis media with effusion at preschool age on some aspects of auditory perception at school age. Ear Hear. 1994 Jun; 15(3):224-31. PMID: 8076720.

71. Hall JW, 3rd, Grose JH, Dev MB, et al. The effect of otitis media with effusion on complex masking tasks in children. Arch Otolaryngol Head Neck Surg. 1998 Aug; 124(8):892-6. PMID: 9708715.

72. ClinicalTrials.gov [Internet]. Identifier: NCT00809601 Randomized Controlled Study of Transmyringeal Tubes. Bethesda (MD): National Library of Medicine (US); 2008. Available at: http://clinicaltrials.gov/ct2/show/NCT00809601. Accessed 2012 May 22.

73. ClinicalTrials.gov [Internet]. Identifier: NCT00546117, A Clinical Trial of Proton Pump Inhibitors to Treat Children With Chronic Otitis Media With Effusion. Bethesda (MD): National Library of Medicine (US); 2007. Available at: http://clinicaltrials.gov/ct2/show/NCT00546117. Accessed 2012 May 22.

74. McCoul ED, Goldstein NA, Koliskor B, et al. A prospective study of the effect of gastroesophageal reflux disease treatment on children with otitis media. Arch Otolaryngol Head Neck Surg. 2011 Jan; 137(1):35-41. PMID: 21242544.

75. van der Pol RJ, Smits MJ, van Wijk MP, et al. Efficacy of proton-pump inhibitors in children with gastroesophageal reflux disease: a systematic review. Pediatrics. 2011 May; 127(5):925-35. PMID: 21464183.

76. ClinicalTrials.gov [Internet]. Identifier: NCT00393900 Middle Ear Pressure Disregulation After Tympanostomy Tube Insertion. Bethesda (MD): National Library of Medicine (US); 2006. Available at: http://clinicaltrials.gov/ct2/show/NCT00393900. Accessed 2012 May 22.

HEALTH EVIDENCE REVIEW COMMISSION (HERC)

DRAFT COVERAGE GUIDANCE: CERVICAL CANCER SCREENING

DATE: XX/XX/XXXX

HERC COVERAGE GUIDANCE

Cervical cancer screening should be covered in women 21 to 30 years old

- Not more than every 3 years for women, with cytology alone
- Co-testing (combined HPV testing and cytology) should not be covered for women under age 30

Cervical cancer screening should be covered in women 30 to 65 years old either with:

- Co-testing not more than every 5 years
- Cytology alone not more than every 3 years

Cervical cancer screening should not be covered for the following populations:

- Women less than age 21
- Women who have had a hysterectomy with removal of cervix for non-cervical cancer related (i.e. high grade precancerous lesion i.e. CIN 2 or 3, or cervical cancer)
- Women over age 65 who have had adequate prior screening and are not otherwise at high risk of cervical cancer
- In women under 30, screening with HPV testing alone or co-testing

Specific testing considerations:

- Liquid based cytology and/or conventional cytology should be covered.
- HPV testing should not be covered for further triaging when low-grade squamous intraepithelial lesions or higher are diagnosed
- The above recommendations also apply to women who have had abnormal testing but whom are indicated to resume routine screening.*

*Management of abnormal cytology and HPV testing is not addressed in this coverage guidance. The United States Preventive Services Task Force (USPSTF) refers to the American Society for Colposcopy and Cervical Pathology 2012 Guideline to address management of abnormal results.

.

DRAFT

Oregon Health Authority

RATIONALE FOR GUIDANCE DEVELOPMENT

The HERC selects topics for guideline development or technology assessment based on the following principles:

- Represents a significant burden of disease
- Represents important uncertainty with regard to efficacy or harms
- Represents important variation or controversy in clinical care
- Represents high costs, significant economic impact
- Topic is of high public interest

Coverage guidance development follows to translate the evidence review to a policy decision. Coverage guidance may be based on an evidence-based guideline developed by the Evidence-based Guideline Subcommittee or a health technology assessment developed by the Heath Technology Assessment Subcommittee. In addition, coverage guidance may utilize an existing evidence report produced by one of HERC's trusted sources, generally within the last three years.

EVIDENCE SOURCES

Hartmann, K.E., Hall, S.A., Nanda, K., et al. (2002). *Screening for cervical cancer* [Internet]. Rockville, MD: Agency for Healthcare Research and Quality (US). Retrieved September 18, 2012, from http://www.ncbi.nlm.nih.gov/books/NBK42831/

Moyer, V.A., & U.S. Preventive Services Task Force. (2012). Screening for cervical cancer: US Preventive Services Task Force recommendation statement. *Annals of Internal Medicine, 156*, 880-891.

Vesco, K.K., Whitlock, E.P., Eder, M., Lin, J., Burda, B.U., Senger, C.A., et al. (2011). *Screening for cervical cancer: A systematic evidence review for the U.S. Preventive Services Task Force.* Evidence Synthesis No. 86. AHRQ Publication No. 11-05156-EF-1. Rockville, MD: Agency for Healthcare Research and Quality.

The summary of evidence in this document is derived directly from these evidence sources, and portions are extracted verbatim.

SUMMARY OF EVIDENCE

Clinical Background

Cervical cancer remains a significant public health issue, even though the incidence and associated mortality of cervical cancer have continued to decrease in the United States since the introduction of cervical cytology screening programs in the 1950s and 60s. In 1950, the Centers for Disease Control (CDC) – Vital Statistics of the United States

DRAFT

reported a death rate of 10.2 per 100,000 for white women, while in 2007 the mortality rate had dropped to 2.2. Incidence varies significantly by age and race/ethnicity.

Cervical cancer does not develop suddenly and is preceded by precancerous changes of the cervix. Precancerous changes of the cervix are histologically defined as cervical intraepithelial neoplasia (CIN) and are identified at varying levels of severity: CIN1, CIN2, and CIN3. The latter includes carcinoma in situ. Progression of neoplasia to invasive cervical cancer is slow. The rate of progression of CIN3 to cancer has recently been estimated as 31.3% in 30 years.

It is well recognized that infection with oncogenic human papilloma virus (HPV) is a necessary, although not sufficient, cause of virtually all cervical cancer. While there are multiple types of HPV, types 16 and 18 alone are responsible for approximately 70% of cervical cancer cases, and HPV is present in 99.7% of cases. The progression from HPV infection to cervical cancer occurs over a series of four steps: 1) HPV transmission, 2) acute HPV infection, 3) persistent HPV infection leading to precancerous changes, and 4) invasive cervical cancer. A high proportion of sexually active women become infected with HPV, but only a small proportion of HPV infections become persistent. Among 4,504 women aged 18 years and older with a cytologic diagnosis of atypical squamous cells of uncertain significance or low-grade squamous intraepithelial lesion, 91% of prevalent HPV infections detected at enrollment cleared within 24 months. These data illustrate that HPV infections are very likely to regress, and persistence of HPV infection is more likely to occur in older women. Numerous analyses, including large cohort studies, have demonstrated that CIN not only progresses, but may also regress. Newer data suggest that CIN1 does not predict any meaningful risk of CIN3.

While it is estimated that around 80% of US women have had cervical cytology screening within the past three years, screening history varies by educational attainment, race/ethnicity, and age. While the great majority of US women have had recent cytology screening, the majority of cervical cancer cases occur in those without such a history.

With regard to screening methods, liquid-based cytology differs from conventional cytology in how the cervical specimen is sent to the cytology laboratory for evaluation. For conventional cytology, the cervical specimen is smeared onto a glass slide immediately after collection and the slide is either sprayed with or placed in fixative. For liquid-based cytology, the sample collected from the cervix is suspended in fixative, then collected by filtration on a membrane, and then transferred onto a microscope slide in a monolayer.

In recent years, high-risk HPV testing has been incorporated into screening and screening triage algorithms, as either a combined test (with cytology, co-test) to determine rescreening interval in women who are cytology negative, or as one possible triage strategy to determine colposcopy. There are many methods available for detecting HPV, including in situ hybridization, polymerase chain reaction, and Hybrid Capture (HC2) technology.

DRAFT

Evidence Review

US Preventive Services Task Force Clinical Considerations

Patient Population under Consideration

This recommendation statement applies to all women who have a cervix, regardless of sexual history. This recommendation statement does not apply to women who have received a diagnosis of a high-grade precancerous cervical lesion or cervical cancer, women with in utero exposure to diethylstilbestrol, or women who are immunocompromised (such as those who are HIV positive).

Screening Tests

The effectiveness of cervical cancer screening observed in the United States over the past several decades is attributed to the use of conventional cytology. Current evidence indicates that there are no clinically important differences between liquid-based cytology and conventional cytology. The USPSTF realizes that the choice of cytology method may not be under the direct control of the clinician and considers cytology screening in appropriate age groups at appropriate intervals to be of substantial net benefit, regardless of method. Human papillomavirus testing with Digene Hybrid Capture 2 (HC2) (Qiagen, Germantown, Maryland) is commonly used in the United States, and both HC2 and polymerase chain reaction– based methods have been evaluated in effectiveness trials. Although alternative HPV detection methods are emerging, the clinical comparability and implications of these methods are not completely understood.

Screening Interval

Screening women aged 21 to 65 years every 3 years with cytology provides a reasonable balance between benefits and harms. Among women aged 30 to 65 years, HPV testing combined with cytology (co-testing) every 5 years offers a comparable balance of benefits and harms and is therefore a reasonable alternative for women in this age group who would prefer to extend the screening interval. Screening with cytology more often than every 3 years confers little additional benefit, with large increases in harms, including additional procedures and assessment and treatment of transient lesions. Treatment of lesions that would otherwise resolve on their own is harmful because it can lead to procedures with unwanted side effects, including the potential for cervical incompetence and preterm labor. Similarly, HPV testing with cytology should not be done more often than every 5 years to maintain a reasonable balance of benefits and harms similar to that seen with cytology alone every 3 years. Among women younger than 30 years, there is adequate evidence that screening with HPV testing (alone or in combination with cytology) confers little to no benefit, and that the harms of HPV testing in this age group are moderate. Therefore, routine screening with HPV in this population is not recommended.

Maintaining the comparability of the benefits and harms of co-testing and cytology alone demands that patients, clinicians, and health care organizations adhere to currently recommended screening intervals, protocols for repeated testing, cytologic thresholds for further diagnostic testing (that is, colposcopy) and treatments, and extended surveillance as recommended by current American Cancer Society/American

DRAFT

Society for Colposcopy and Cervical Pathology/American Society for Clinical Pathology (ACS/ASCCP/ASCP) guidelines. Women who choose co-testing to increase their screening interval (and potentially decrease testing) should be aware that positive screening results are more likely with HPV-based strategies than with cytology alone and that some women may require prolonged surveillance with additional frequent testing if they have persistently positive HPV results. Because HPV test results may be positive among women who would otherwise be advised to end screening at age 65 years on the basis of previously normal cytology results alone, the likelihood of continued testing may increase with HPV testing. The percentage of US women undergoing co-testing who will have a normal cytology test result and a positive HPV test result (and who will therefore require additional testing) ranges from 11% among women aged 30 to 34 years to 2.6% among women aged 60 to 65 years.

Triage of Women with Atypical Squamous Cells of Uncertain Significance
For the triage of women with atypical squamous cells of uncertain significance cytology to colposcopy, a single HC2 test has a higher sensitivity and similar specificity compared to single repeat cytology at a threshold of atypical squamous cells of uncertain significance for the detection of CIN2+. No additional benefit occurs when HC2 triage is combined with cytology, but this strategy increases false positives. The HC2 does not appear useful for the triage of women with low-grade squamous intraepithelial lesion cytology because such a high proportion of women will test positive. Human papilloma virus testing has few unique harms compared with cytology screening, but a positive HPV test may increase anxiety and distress, in the short-term only.

Timing of Screening
Women Younger Than Age 21 Years
Cervical cancer is rare before age 21 years. The USPSTF found little evidence to determine whether and how sexual history should affect the age at which to begin screening. Although exposure of cervical cells to sexually transmitted HPV during vaginal intercourse may lead to cervical carcinogenesis, the process has multiple steps, involves regression, and is generally not rapid. There is evidence that screening earlier than age 21 years, regardless of sexual history, would lead to more harm than benefit. The harms are greater in this younger age group because abnormal test results are likely to be transient and to resolve on their own; in addition, treatment may have an adverse effect on childbearing.

Women Older Than Age 65 Years
Clinicians and patients should base the decision to end screening on whether the patient meets the criteria for adequate prior testing and appropriate follow-up per established guidelines. The ACS/ASCCP/ASCP guidelines define adequate prior screening as 3 consecutive negative cytology results or 2 consecutive negative HPV results within 10 years before cessation of screening, with the most recent test occurring within 5 years. They further state that routine screening should continue for at least 20 years after spontaneous regression or appropriate management of a high-grade precancerous lesion, even if this extends screening past age 65 years. The ACS

DRAFT

further states that screening should not resume after cessation in women older than age 65 years, even if a woman reports having a new sexual partner.

Women Older Than Age 65 Years Who Have Never Been Screened
Screening may be clinically indicated in older women for whom the adequacy of prior screening cannot be accurately accessed or documented. Women with limited access to care, minority women, and women from countries where screening is not available may be less likely to meet the criteria for adequate prior screening. The USPSTF realizes that certain considerations may support screening in women older than age 65 years who are otherwise considered high risk (such as women with a high grade precancerous lesion or cervical cancer, women with in utero exposure to diethylstilbestrol, or women who are immunocompromised).

Assessment of Risk
It is well-established that HPV infection is associated with nearly all cases of cervical cancer. Other risk factors include HIV infection, a compromised immune system, in utero exposure to diethylstilbestrol, and previous treatment of a high-grade precancerous lesion or cervical cancer. Women who have had a hysterectomy with removal of the cervix and who do not have a history of a high-grade precancerous lesion or cervical cancer are not at risk for cervical cancer and should not be screened. Women who had their cervix removed during surgery for ovarian or endometrial cancer are not at high risk for cervical cancer and would not benefit from screening. Clinicians should confirm through review of surgical records or direct examination that the cervix was removed.

Recommendations

These recommendations apply to women who have a cervix, regardless of sexual history. These recommendations do not apply to women who have received a diagnosis of a high-grade precancerous cervical lesion or cervical cancer, women with in utero exposure to diethylstilbestrol, or women who are immunocompromised (such as those who are HIV positive).

- The USPSTF recommends screening for cervical cancer in women ages 21 to 65 years with cytology (Pap smear) every 3 years or, for women ages 30 to 65 years who want to lengthen the screening interval, screening with a combination of cytology and human papillomavirus (HPV) testing every 5 years. Grade: A Recommendation.

- The USPSTF recommends against screening for cervical cancer in women younger than age 21 years. Grade: D Recommendation.

- The USPSTF recommends against screening for cervical cancer in women older than age 65 years who have had adequate prior screening and are not otherwise at high risk for cervical cancer. Grade: D Recommendation.

- The USPSTF recommends against screening for cervical cancer in women who have had a hysterectomy with removal of the cervix and who do not have a

history of a high-grade precancerous lesion (i.e., cervical intraepithelial neoplasia [CIN] grade 2 or 3) or cervical cancer. Grade: D Recommendation.

- The USPSTF recommends against screening for cervical cancer with HPV testing, alone or in combination with cytology, in women younger than age 30 years. Grade: D Recommendation.

Overall Summary

A reasonable age at which to initiate cervical cancer screening in women is age 21. For cytology-based screening, liquid-based cytology does not differ from conventional cytology in sensitivity, specificity, or relative CIN detection. Screening women aged 21 to 65 years every 3 years with cytology provides a reasonable balance between benefits and harms. Among women aged 30 to 65 years, HPV testing combined with cytology (co-testing) every 5 years offers a comparable balance of benefits and harms. Screening with cytology more often than every 3 years confers little additional benefit, with large increases in harms. Among women younger than 30 years, screening with HPV testing (alone or in combination with cytology) confers little to no benefit but has moderate harms. Treatment of lesions that would otherwise resolve on their own is harmful because it can lead to procedures with unwanted side effects, including the potential for cervical incompetence and preterm labor. For the triage of women with atypical squamous cells of uncertain significance cytology to colposcopy, a single HC2 test has a higher sensitivity and similar specificity compared to single repeat cytology, but there are no additional benefits when HC2 triage is combined with cytology. The HC2 is not useful for the triage of women with low-grade squamous intraepithelial lesion cytology. It is reasonable to discontinue routine cervical cancer screening for women older than age 65 years who have had adequate screening with negative results and who are not otherwise at high risk for cervical cancer, and for women who have undergone a hysterectomy in which the cervix was removed, unless it was performed because of cervical cancer.

PROCEDURE

Pap smear
HPV testing

DIAGNOSES

Cervical cancer screening

APPLICABLE CODES

CODES	DESCRIPTION
ICD-9 Diagnosis Codes	
V76.2	Special screening for malignant neoplasms; cervix
V73.81	Special screening for viral and chlamydial diseases; human papilloma virus
079.4	Viral and chlamydial infection in conditions classified elsewhere; HPV
795.0	Abnormal PAP smear of cervix and cervical HPV

DRAFT

CODES	DESCRIPTION
ICD-9 Volume 3 (Procedure Codes)	
None	
CPT Codes	
88141	Cytopathology, cervical or vaginal, requiring interpretation by physician
88142-3	Cytopathology, cervical or vaginal, collected in preservative fluid, manual screening
88147-8	Cytopathology smears, cervical or vaginal
88150-4	Cytopathology slides, cervical or vaginal
88164-7	Cytopathology slides, cervical or vaginal, Bethesda system
88174-5	Cytopathology, cervical or vaginal, collected in preservative fluid, automated screening
87620	Detection infectious agent by probe technique; HPV, direct
87621	Detection infectious agent by probe technique; HPV, amplified
HCPCS Codes	
G0123-4	Screening cytopathology, cervical or vaginal, collected in preservative fluid, automated thin-layer prep
G0141	Screening cytopathology, cervical or vaginal, requiring interpretation by physician
G0143-5	Screening cytopathology, cervical or vaginal, collected in preservative fluid, automated thin-layer prep
G0147-8	Screening cytopathology smears, cervical or vaginal

Note: Inclusion on this list does not guarantee coverage

Coverage guidance is prepared by the Health Evidence Review Commission (HERC), HERC staff, and subcommittee members. The evidence summary is prepared by the Center for Evidence-based Policy at Oregon Health & Science University (the Center). This document is intended to guide public and private purchasers in Oregon in making informed decisions about health care services.

The Center is not engaged in rendering any clinical, legal, business or other professional advice. The statements in this document do not represent official policy positions of the Center. Researchers involved in preparing this document have no affiliations or financial involvement that conflict with material presented in this document.

DRAFT

Screening for Cervical Cancer: U.S. Preventive Services Task Force Recommendation Statement

Virginia A. Moyer, MD, MPH, on behalf of the U.S. Preventive Services Task Force*

Description: Update of the 2003 U.S. Preventive Services Task Force (USPSTF) recommendation statement on screening for cervical cancer.

Methods: The USPSTF reviewed new evidence on the comparative test performance of liquid-based cytology and the benefits and harms of human papillomavirus (HPV) testing as a stand-alone test or in combination with cytology. In addition to the systematic evidence review, the USPSTF commissioned a decision analysis to help clarify the age at which to begin and end screening, the optimal interval for screening, and the relative benefits and harms of different strategies for screening (such as cytology and co-testing).

Recommendations: This recommendation statement applies to women who have a cervix, regardless of sexual history. This recommendation statement does not apply to women who have received a diagnosis of a high-grade precancerous cervical lesion or cervical cancer, women with in utero exposure to diethylstilbestrol, or women who are immunocompromised (such as those who are HIV positive).

The USPSTF recommends screening for cervical cancer in women aged 21 to 65 years with cytology (Papanicolaou smear) every 3 years or, for women aged 30 to 65 years who want to lengthen the screening interval, screening with a combination of cytology and HPV testing every 5 years. See the Clinical Considerations for discussion of cytology method, HPV testing, and screening interval (A recommendation).

The USPSTF recommends against screening for cervical cancer in women younger than age 21 years (D recommendation).

The USPSTF recommends against screening for cervical cancer in women older than age 65 years who have had adequate prior screening and are not otherwise at high risk for cervical cancer. See the Clinical Considerations for discussion of adequacy of prior screening and risk factors (D recommendation).

The USPSTF recommends against screening for cervical cancer in women who have had a hysterectomy with removal of the cervix and who do not have a history of a high-grade precancerous lesion (cervical intraepithelial neoplasia grade 2 or 3) or cervical cancer (D recommendation).

The USPSTF recommends against screening for cervical cancer with HPV testing, alone or in combination with cytology, in women younger than age 30 years (D recommendation).

Ann Intern Med. 2012;156:880-891. www.annals.org

For author affiliation, see end of text.

* For a list of the members of the USPSTF, see the **Appendix** (available at www.annals.org).

This article was published at www.annals.org on 15 March 2012.

The U.S. Preventive Services Task Force (USPSTF) makes recommendations about the effectiveness of specific clinical preventive services for patients without related signs or symptoms.

It bases its recommendations on the evidence of both the benefits and harms of the service, and an assessment of the balance. The USPSTF does not consider the costs of providing a service in this assessment.

The USPSTF recognizes that clinical decisions involve more considerations than evidence alone. Clinicians should understand the evidence but individualize decision making to the specific patient or situation. Similarly, the USPSTF notes that policy and coverage decisions involve considerations in addition to the evidence of clinical benefits and harms.

See also:

Print

Editorial comment. 896

Summary for Patients. I-44

Web-Only

Appendix

CME quiz (preview on page I-23)

Conversion of graphics into slides

SUMMARY OF RECOMMENDATIONS AND EVIDENCE

This recommendation statement applies to women who have a cervix, regardless of sexual history. This recommendation statement does not apply to women who have received a diagnosis of a high-grade precancerous cervical lesion or cervical cancer, women with in utero exposure to diethylstilbestrol, or women who are immunocompromised (such as those who are HIV positive).

The USPSTF recommends screening for cervical cancer in women aged 21 to 65 years with cytology (Papanicolaou [Pap] smear) every 3 years or, for women aged 30 to 65 years who want to lengthen the screening interval,

screening with a combination of cytology and human papillomavirus (HPV) testing every 5 years. See the Clinical Considerations for discussion of cytology method, HPV testing, and screening interval (A recommendation).

The USPSTF recommends against screening for cervical cancer in women younger than age 21 years (D recommendation).

The USPSTF recommends against screening for cervical cancer in women older than age 65 years who have had adequate prior screening and are not otherwise at high risk for cervical cancer. See the Clinical Considerations for discussion of adequacy of prior screening and risk factors (D recommendation).

The USPSTF recommends against screening for cervical cancer in women who have had a hysterectomy with removal of the cervix and who do not have a history of a high-grade precancerous lesion (cervical intraepithelial neoplasia [CIN] grade 2 or 3) or cervical cancer (D recommendation).

The USPSTF recommends against screening for cervical cancer with HPV testing, alone or in combination with cytology, in women younger than age 30 years (D recommendation).

See the **Figure** for a summary of the recommendations and suggestions for clinical practice.

Table 1 describes the USPSTF grades, and **Table 2** describes the USPSTF classification of levels of certainty about net benefit.

Rationale

Importance

The age-adjusted annual incidence rate of cervical cancer is 6.6 cases per 100 000 women, according to data from 2008 (1–3). An estimated 12 200 new cases of cervical cancer and 4210 deaths occurred in the United States in 2010 (1). Cervical cancer deaths in the United States have decreased dramatically since the implementation of widespread cervical cancer screening. Most cases of cervical cancer occur in women who have not been appropriately screened (2, 3). Strategies that aim to ensure that all women are screened at the appropriate interval and receive adequate follow-up are most likely to be successful in further reducing cervical cancer incidence and mortality in the United States.

Detection

Screening with cervical cytology or testing for multiple oncogenic HPV types (a test for the presence of >2 high-risk or carcinogenic HPV types, hereafter called *HPV testing*) can lead to detection of high-grade precancerous cervical lesions and cervical cancer.

Benefits of Detection and Early Intervention and Treatment

Women Aged 21 to 65 Years

There is convincing evidence that screening women aged 21 to 65 years with cytology every 3 years substantially reduces cervical cancer incidence and mortality. Among women aged 30 to 65 years, there is adequate evidence that screening with a combination of cytology and HPV testing (co-testing) every 5 years provides benefits similar to those seen with cytology screening alone every 3 years.

Among women younger than age 30 years, there is adequate evidence that screening with HPV testing (alone or in combination with cytology) confers little to no benefit.

Women Younger Than Age 21 Years

There is adequate evidence that screening women younger than age 21 years (regardless of sexual history) does not reduce cervical cancer incidence and mortality compared with beginning screening at age 21 years (4).

Women Older Than Age 65 Years

There is adequate evidence that screening women older than age 65 years who have had adequate prior screening and are not otherwise at high risk provides little to no benefits.

Women After Hysterectomy

There is convincing evidence that continued screening after hysterectomy with removal of the cervix for indications other than a high-grade precancerous lesion or cervical cancer provides no benefits.

Harms of Detection and Early Intervention and Treatment

Screening with cervical cytology or HPV testing can lead to harms, and the harms of screening can take many forms. Abnormal test results can lead to more frequent testing and invasive diagnostic procedures, such as colposcopy and cervical biopsy. Evidence from randomized, controlled trials and observational studies indicates that harms from these diagnostic procedures include vaginal bleeding, pain, infection, and failure to diagnose (due to inadequate sampling). Abnormal screening test results are also associated with mild psychological harms; short-term increases in anxiety, distress, and concern about health have been reported with cytology and HPV testing.

Harms of Treatment of Screening-Detected Disease

The harms of treatment include risks from the treatment procedure itself and the potential downstream consequences of treatment. Summary evidence from observational studies indicates that some treatments for precancerous lesions (such as cold-knife conization and loop excision) are associated with adverse pregnancy outcomes, such as preterm delivery, that can lead to low birthweight in infants and perinatal death (2). Evidence is convincing that many precancerous cervical lesions will re-

Figure. **Screening for cervical cancer: clinical summary of U.S. Preventive Services Task Force recommendation.**

Annals of Internal Medicine

U.S. Preventive Services TASK FORCE
www.USPreventiveServicesTaskForce.org

SCREENING FOR CERVICAL CANCER

CLINICAL SUMMARY OF U.S. PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

Population	Women Aged 21 to 65 Years	Women Aged 30 to 65 Years	Women Younger Than Age 21 Years	Women Older Than Age 65 Years Who Have Had Adequate Prior Screening and Are Not High Risk	Women After Hysterectomy With Removal of the Cervix With No History of High-Grade Precancer or Cervical Cancer	Women Younger Than Age 30 Years
Recommendation	Screen with cytology (Pap smear) every 3 years Grade: A	Screen with cytology every 3 years or co-testing (cytology/ human papillomavirus testing [HPV]) every 5 years Grade: A	Do not screen Grade: D	Do not screen Grade: D	Do not screen Grade: D	Do not screen with HPV testing (alone or with cytology) Grade: D

Risk Assessment	HPV infection is associated with nearly all cases of cervical cancer. Other factors that increase a woman's risk for cervical cancer include HIV infection, a compromised immune system, in utero exposure to diethylstilbestrol, and previous treatment of a high-grade precancerous lesion or cervical cancer.					
Screening Tests and Interval	Screening women aged 21 to 65 years every 3 years with cytology provides a reasonable balance between benefits and harms. Screening with cytology more often than every 3 years confers little additional benefit, with large increases in harms. HPV testing combined with cytology (co-testing) every 5 years in women aged 30 to 65 years offers a comparable balance of benefits and harms, and is therefore a reasonable alternative for women in this age group who would prefer to extend the screening interval.					
Timing of Screening	Screening women younger than age 21 years, regardless of sexual history, leads to more harms than benefits. Clinicians and patients should base the decision to end screening on whether the patient meets the criteria for adequate prior testing and appropriate follow-up, per established guidelines.					
Interventions	Screening aims to identify high-grade precancerous cervical lesions to prevent development of cervical cancer and early-stage asymptomatic invasive cervical cancer. High-grade lesions may be treated with ablative and excisional therapies, including cryotherapy, laser ablation, loop excision, and cold-knife conization. Early-stage cervical cancer may be treated with surgery (hysterectomy) or chemoradiation.					
Balance of Benefits and Harms	The benefits of screening with cytology every 3 years substantially outweigh the harms.	The benefits of screening with co-testing (cytology/HPV testing) every 5 years outweigh the harms.	The harms of screening earlier than age 21 years outweigh the benefits.	The benefits of screening after age 65 years do not outweigh the potential harms.	The harms of screening after hysterectomy outweigh the benefits.	The potential harms of screening with HPV testing (alone or with cytology) outweigh the potential benefits.
Other Relevant USPSTF Recommendations	The USPSTF has made recommendations on screening for breast cancer and ovarian cancer, as well as genetic risk assessment and BRCA mutation testing for breast and ovarian cancer susceptibility. These recommendations are available at www.uspreventiveservicestaskforce.org.					

For a summary of the evidence systematically reviewed in making this recommendation, the full recommendation statement, and supporting documents, please go to www.uspreventiveservicestaskforce.org.

Pap = Papanicolaou.

gress and that other lesions are so indolent and slow-growing that they will not become clinically important over a woman's lifetime; identification and treatment of these lesions constitute overdiagnosis. It is difficult to estimate the precise magnitude of overdiagnosis associated with any screening or treatment strategy, but it is of concern because it confers no benefit and leads to unnecessary surveillance, diagnostic tests, and treatments with the associated harms.

Women Aged 21 to 65 Years

There is adequate evidence that the harms of screening for cervical cancer with cytology alone or in combination with HPV testing in women aged 30 to 65 years are moderate. Positive screening results are more common with strategies that include HPV testing than with strategies that use cytology alone. Therefore, the likelihood of prolonged surveillance and overtreatment may increase with strategies that incorporate HPV testing. Cervical

Table 1. **What the USPSTF Grades Mean and Suggestions for Practice**

Grade	Definition	Suggestions for Practice
A	The USPSTF recommends the service. There is high certainty that the net benefit is substantial.	Offer or provide this service.
B	The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial.	Offer or provide this service.
C	*Note: The following statement is undergoing revision.* Clinicians may provide this service to selected patients depending on individual circumstances. However, for most individuals without signs or symptoms there is likely to be only a small benefit from this service.	Offer or provide this service only if other considerations support the offering or providing the service in an individual patient.
D	The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits.	Discourage the use of this service.
I Statement	The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.	Read the clinical considerations section of USPSTF Recommendation Statement. If the service is offered, patients should understand the uncertainty about the balance of benefits and harms.

treatments may increase the risk for adverse pregnancy outcomes (for example, cervical insufficiency and preterm delivery) in women who have not yet completed childbearing.

Women Younger Than Age 30 Years

There is adequate evidence that the harms of HPV testing (alone or in combination with cytology) in women younger than age 30 years are moderate.

Women Younger Than Age 21 Years

There is adequate evidence that the harms of screening in women younger than age 21 years are moderate.

Women Older Than Age 65 Years

There is adequate evidence that the harms of screening in women older than age 65 years who have had adequate prior screening and are not otherwise at high risk are at least small.

Women After Hysterectomy

There is adequate evidence that screening after hysterectomy among women who do not have a history of a high-grade precancerous lesion or cervical cancer is associated with harms.

USPSTF Assessment

The USPSTF concludes that for women aged 21 to 65 years, there is high certainty that the benefits of screening with cytology every 3 years substantially outweigh the harms. For women aged 30 to 65 years, there is high certainty that the benefits of screening with a combination of cytology and HPV testing (co-testing) every 5 years outweigh the harms.

For women younger than age 21 years, regardless of sexual history, there is moderate certainty that the harms of screening outweigh the benefits.

For women older than age 65 years who have had adequate prior screening and are not otherwise at high risk for cervical cancer, there is moderate certainty that the benefits of screening do not outweigh the potential harms.

For women who have had a hysterectomy with removal of the cervix for indications other than a high-grade

Table 2. **Levels of Certainty Regarding Net Benefit***

Level of Certainty	Description
High	The available evidence usually includes consistent results from well-designed, well-conducted studies in representative primary care populations. These studies assess the effects of the preventive service on health outcomes. This conclusion is therefore unlikely to be strongly affected by the results of future studies.
Moderate	The available evidence is sufficient to determine the effects of the preventive service on health outcomes, but confidence in the estimate is constrained by such factors as: the number, size, or quality of individual studies; inconsistency of findings across individual studies; limited generalizability of findings to routine primary care practice; and lack of coherence in the chain of evidence. As more information becomes available, the magnitude or direction of the observed effect could change, and this change may be large enough to alter the conclusion.
Low	The available evidence is insufficient to assess effects on health outcomes. Evidence is insufficient because of: the limited number or size of studies; important flaws in study design or methods; inconsistency of findings across individual studies; gaps in the chain of evidence; findings that are not generalizable to routine primary care practice; and a lack of information on important health outcomes. More information may allow estimation of effects on health outcomes.

* The USPSTF defines *certainty* as "likelihood that the USPSTF assessment of the net benefit of a preventive service is correct." The net benefit is defined as benefit minus harm of the preventive service as implemented in a general, primary care population. The USPSTF assigns a certainty level based on the nature of the overall evidence available to assess the net benefit of a preventive service.

precancerous lesion or cancer, there is high certainty that the harms of screening outweigh the benefits.

For women younger than age 30 years, there is moderate certainty that the potential harms of screening with HPV testing (alone or in combination with cytology) outweigh the potential benefits.

Clinical Considerations

Patient Population Under Consideration

This recommendation statement applies to all women who have a cervix, regardless of sexual history. This recommendation statement does not apply to women who have received a diagnosis of a high-grade precancerous cervical lesion or cervical cancer, women with in utero exposure to diethylstilbestrol, or women who are immunocompromised (such as those who are HIV positive).

Screening Tests

The effectiveness of cervical cancer screening observed in the United States over the past several decades is attributed to the use of conventional cytology. Current evidence indicates that there are no clinically important differences between liquid-based cytology and conventional cytology. The USPSTF realizes that the choice of cytology method may not be under the direct control of the clinician and considers cytology screening in appropriate age groups at appropriate intervals to be of substantial net benefit, regardless of method. Human papillomavirus testing with Digene Hybrid Capture 2 (HC2) (Qiagen, Germantown, Maryland) is commonly used in the United States, and both HC2 and polymerase chain reaction–based methods have been evaluated in effectiveness trials. Although alternative HPV detection methods are emerging, the clinical comparability and implications of these methods are not completely understood.

Screening Interval

Screening women aged 21 to 65 years every 3 years with cytology provides a reasonable balance between benefits and harms. Among women aged 30 to 65 years, HPV testing combined with cytology (co-testing) every 5 years offers a comparable balance of benefits and harms and is therefore a reasonable alternative for women in this age group who would prefer to extend the screening interval. Screening with cytology more often than every 3 years confers little additional benefit, with large increases in harms, including additional procedures and assessment and treatment of transient lesions. Treatment of lesions that would otherwise resolve on their own is harmful because it can lead to procedures with unwanted side effects, including the potential for cervical incompetence and preterm labor. Similarly, HPV testing with cytology should not be done more often than every 5 years to maintain a reasonable balance of benefits and harms similar to that seen with cytology alone every 3 years. Maintaining the comparability of the benefits and harms of co-testing and cytology alone demands that patients, clinicians, and health care organizations adhere to currently recommended screening intervals, protocols for repeated testing, cytologic thresholds for further diagnostic testing (that is, colposcopy) and treatments, and extended surveillance as recommended by current American Cancer Society/American Society for Colposcopy and Cervical Pathology/American Society for Clinical Pathology (ACS/ASCCP/ASCP) guidelines.

Women who choose co-testing to increase their screening interval (and potentially decrease testing) should be aware that positive screening results are more likely with HPV-based strategies than with cytology alone and that some women may require prolonged surveillance with additional frequent testing if they have persistently positive HPV results. Because HPV test results may be positive among women who would otherwise be advised to end screening at age 65 years on the basis of previously normal cytology results alone, the likelihood of continued testing may increase with HPV testing. The percentage of U.S. women undergoing co-testing who will have a normal cytology test result and a positive HPV test result (and who will therefore require additional testing) ranges from 11% among women aged 30 to 34 years to 2.6% among women aged 60 to 65 years (5, 6).

Timing of Screening

Women Younger Than Age 21 Years

Cervical cancer is rare before age 21 years. The USPSTF found little evidence to determine whether and how sexual history should affect the age at which to begin screening. Although exposure of cervical cells to sexually transmitted HPV during vaginal intercourse may lead to cervical carcinogenesis, the process has multiple steps, involves regression, and is generally not rapid. There is evidence that screening earlier than age 21 years, regardless of sexual history, would lead to more harm than benefit (4). The harms are greater in this younger age group because abnormal test results are likely to be transient and to resolve on their own; in addition, treatment may have an adverse effect on childbearing.

Women Older Than Age 65 Years

Clinicians and patients should base the decision to end screening on whether the patient meets the criteria for adequate prior testing and appropriate follow-up per established guidelines. The ACS/ASCCP/ASCP guidelines define adequate prior screening as 3 consecutive negative cytology results or 2 consecutive negative HPV results within 10 years before cessation of screening, with the most recent test occurring within 5 years. They further state that routine screening should continue for at least 20 years after spontaneous regression or appropriate management of a high-grade precancerous lesion, even if this extends screening past age 65 years (7). The ACS further states that screening should not resume after cessation in

women older than age 65 years, even if a woman reports having a new sexual partner.

Women Older Than Age 65 Years Who Have Never Been Screened

Screening may be clinically indicated in older women for whom the adequacy of prior screening cannot be accurately accessed or documented. Women with limited access to care, minority women, and women from countries where screening is not available may be less likely to meet the criteria for adequate prior screening. The USPSTF realizes that certain considerations may support screening in women older than age 65 years who are otherwise considered high risk (such as women with a high-grade precancerous lesion or cervical cancer, women with in utero exposure to diethylstilbestrol, or women who are immunocompromised).

Assessment of Risk

It is well-established that HPV infection is associated with nearly all cases of cervical cancer. Other risk factors include HIV infection, a compromised immune system, in utero exposure to diethylstilbestrol, and previous treatment of a high-grade precancerous lesion or cervical cancer.

Women who have had a hysterectomy with removal of the cervix and who do not have a history of a high-grade precancerous lesion or cervical cancer are not at risk for cervical cancer and should not be screened. Women who had their cervix removed during surgery for ovarian or endometrial cancer are not at high risk for cervical cancer and would not benefit from screening. Clinicians should confirm through review of surgical records or direct examination that the cervix was removed.

Treatment

Screening aims to identify high-grade precancerous cervical lesions to prevent development of cervical cancer and early-stage asymptomatic invasive cervical cancer. High-grade lesions may be treated with ablative and excisional therapies, including cryotherapy, laser ablation, loop excision, and cold-knife conization. Early-stage cervical cancer may be treated with surgery (hysterectomy) or chemoradiation. The treatment of precancerous rather than early-stage cancerous lesions is unique to cervical cancer and is the foundation of the success of cervical cancer screening. Treatment of precancerous lesions is less invasive than treatment of cancer and results in fewer adverse effects.

Other Approaches to Prevention

Many individuals and clinicians have used the annual Pap smear screening visit as an opportunity to discuss other health problems and preventive measures. Individuals, clinicians, and health systems should seek effective ways to facilitate the receipt of recommended preventive services at intervals that are beneficial to the patient. Efforts should also be made to ensure that individuals are able to seek care for additional health concerns as they present.

The overall effect of HPV vaccination on high-grade precancerous cervical lesions and cervical cancer is not yet known. Current trials do not provide data on long-term efficacy (8); therefore, the possibility that vaccination might reduce the need for screening with cytology alone or in combination with HPV testing is not established. Given these uncertainties, women who have been vaccinated should continue to be screened.

Other Considerations

Research Needs and Gaps

There are notable limitations to the current evidence. There is limited direct evidence on the harms of various screening strategies that incorporate HPV testing. Additional data from ongoing trials on cervical cancer outcomes and the resulting number of false-positive test results, colposcopies, and biopsies should help to clarify some of the current uncertainties related to strategies that include HPV testing. Moreover, these data should help to better assess the comparative effectiveness and harms of various screening strategies using cytology and HPV testing alone, in combination, or sequentially.

An important clinical limitation of the current evidence is the lack of long-term cumulative data from screening trials on cervical cancer. Much of the data to date are limited to detection of CIN grade 3. Although CIN3 may be considered an acceptable surrogate for cancer, additional data are needed to determine benefits, harms, and net benefit. Future screening trials should plan for and report round-specific data as well as cumulative results from multiple screening rounds to obtain useful cumulative data on CIN3, cervical cancer by stage and type, and program requirements (such as colposcopy, biopsy, treatments, or harms of treatment). More complete outcomes data will help to better assess the relative benefits of different screening strategies, particularly in comparing various approaches involving cytology and HPV testing.

There is limited evidence on the benefits and harms of HPV testing alone as a screening strategy. An emerging chain of evidence suggests that HPV testing followed by cytology in women with positive HPV test results may also be a reasonable screening strategy. Ongoing studies, such as the HPV FOCAL (HPV Testing for Cervical Cancer Screening) trial, which compares HPV with cytology triage to cytology with HPV triage of test results interpreted as atypical squamous cells of undetermined significance, should provide relevant direct evidence on HPV testing that applies to current U.S. practice.

Finally, more research is needed to determine whether and how individual risk factors may be used to tailor screening, thereby preventing overdiagnosis and overutilization of resources in women at low risk for cervical cancer, as well as underdiagnosis in those at high risk. Deter-

mining risk factors that move lower-risk women (such as older women with normal cytology findings or negative HPV test results) into higher risk categories (such as older women with positive HPV and negative cytology results or exposure to new partners) will also be important.

Discussion

Burden of Disease

Cervical cancer incidence and mortality have declined in the United States since the introduction of cervical cancer screening in the 1950s and 1960s. The current annual incidence rate is 6.6 cases per 100 000 women, and the age-adjusted mortality rate is 2.4 deaths per 100 000 (for 2003 to 2007) (1). However, cervical cancer still remains a substantial public health issue. Incidence rates (2004 to 2008) vary by age and race or ethnicity; Hispanic (11.1 per 100 000) and black (10.0 per 100 000) women experience the highest rates, whereas non-Hispanic white (7.4 per 100 000), American Indian and Alaska Native (7.8 per 100 000), and Asian and Pacific Islander (7.3 per 100 000) women have lower rates (1). Cervical cancer most commonly occurs in women aged 35 to 55 years. In contrast to cervical cancer, abnormal cytology test results and precancerous lesions are fairly common. According to Centers for Disease Control and Prevention data from low-income, uninsured, and underserved women, approximately 3.0% of cytology test results are abnormal (9).

Scope of Review

To update the 2003 recommendation, the USPSTF commissioned a targeted systematic review of the evidence on screening strategies incorporating HPV testing that may apply to current screening practices in the United States. The USPSTF reviewed new evidence regarding the comparative test performance of liquid-based cytology and the benefits and harms of HPV testing as a standalone test or in combination with cytology.

In addition to the systematic evidence review, the USPSTF commissioned a decision analysis to help clarify the age at which to begin and end screening, the optimal interval for screening, and the relative benefits and harms of different strategies for screening (such as cytology and co-testing). The USPSTF uses modeling as a complement to evidence reported in the systematic review, to provide information about alternate screening or treatment strategies in the absence of direct evidence, or when the complexities required to conduct randomized, controlled trials to address knowledge gaps would preclude the ability to obtain direct evidence. The USPSTF does not use modeling to make recommendations for or against screening or treatment.

The USPSTF did not review evidence on automated screening technologies because they are less relevant to primary care clinicians, and it did not review evidence on HPV vaccination because data to determine long-term vaccine efficacy or how vaccination will affect screening are limited.

Accuracy of Screening Tests

Liquid-Based Cytology Compared With Conventional Cytology

Evidence suggests that there are no clinically meaningful differences in accuracy between liquid-based cytology and conventional cytology. One large, good-quality randomized trial (10) and one large, fair-quality randomized trial (11) of more than 130 000 women compared the 2 screening methods and found no difference in detection of CIN2+ or CIN3+ at any cytologic threshold of positivity.

HPV Testing Compared With Cytology

Evidence from good- and fair-quality observational studies indicates that HPV testing generally has a higher sensitivity but lower specificity (that is, more false-positive test results) than does cytology in the detection of CIN2+ and CIN3+ (12–18). False-positive rates are higher among women younger than age 30 to 35 years than women in older age groups because of the higher prevalence of HPV, but the incidence of cervical cancer is lower in the former age group.

Effectiveness of Early Detection or Treatment

Introduction of screening to populations naive to screening reduces cervical cancer rates by 60% to 90% within 3 years of implementation (19). The reduction of mortality and morbidity associated with the introduction of cytology-based screening is consistent and equally dramatic across populations. Correlational studies of cervical cancer trends in countries in North America and Europe demonstrate dramatic reductions in incidence of invasive cervical cancer and a 20% to 60% reduction in cervical cancer mortality since the onset of widespread screening.

No published studies have evaluated, in an ideal way, the age at which to begin screening, the age at which to end screening, and how often to screen. The USPSTF considered the following types of evidence to determine when screening for cervical cancer should begin: incidence, prevalence, and mortality of cervical cancer in young women; the natural history of precancerous lesions and HPV infection; and the effects of screening in populations of young women. Cervical cancer in women younger than age 20 years is rare; according to U.S. Surveillance, Epidemiology, and End Results (SEER) data, 0.1% of all incident cancer cases occur in women younger than age 20 years. Older data from SEER (1) report declining rates of cervical cancer in the years 1973 to 1999; in some years, no cases occurred in women younger than age 20 years (2, 3, 18). Deaths due to cervical cancer in women younger than age 20 years are also rare; fewer than 16 such deaths occurred in the United States from 1992 through 2008 (1). Precancerous lesions are also uncommon. Prevalence of CIN3 among women younger than age 20 years is estimated at 0.2% (20, 21), with a concurrent rate of false-positive cytology results of about 3.1% (21). Because of the lack of

direct evidence, the USPSTF considered results of decision analyses using the best available data to estimate and understand the benefits and harms of screening at different starting ages and intervals; colposcopy was used as a proxy measure for harms. Results of the analyses show that screening every 3 years with cytology starting at age 21 years confers a similar number of life-years as does annual screening (69 247 vs. 69 213 per 1000 women), yet prompts fewer than half of the number of colposcopies and fewer false-positive test results. Varying the age at which to start screening shows no benefit to starting earlier than age 21 years; screening with cytology every 3 years starting at 15 years of age, 18 years of age, and 21 years of age finds cervical cancer death rates of 1.54, 1.54, and 1.55 per 1000 women, respectively (4). The results of these analyses suggest that screening beginning at age 21 years with an interval of 3 years provides the most acceptable balance of benefits and harms.

When deliberating on the age at which to end screening, the USPSTF considered the incidence of cervical cancer in older women and whether there is a difference in the pattern of cervical cancer incidence in screened versus unscreened women. The incidence and prevalence of CIN peak in the midreproductive years and begin to decline in approximately the fourth decade of life, a general pattern also apparent among some previously unscreened women. Cervical cancer in older women is not more aggressive or rapidly progressive than that in younger women. Finally, the rate of high-grade squamous intraepithelial lesions diagnosed by cytology is low among older women who have been previously screened. Modeling studies of the age at which to end screening indicate no substantial benefit beyond age 65 years in women who have been previously screened. Specifically, varying the age at which to end screening from 65 years to 95 years by 5-year intervals provides a very small (<1 life-year) improvement in life-years after age 65 years but increases potential harms due to false-positive results and the increase in the number of colposcopies and cervical biopsies (4).

Although screening women older than age 65 years who have an adequate screening history is not recommended, modeling studies suggest that screening women who have never been screened would reduce mortality by 74% (3, 22). Strategies that include screening previously unscreened women every 2 to 5 years and ending at age 70 to 75 years represent reasonable tradeoffs between benefits and harms (4). Current guidelines define adequate screening as 3 consecutive negative cytology results or 2 consecutive negative HPV results within 10 years before cessation of screening, with the most recent test performed within 5 years (7). Women with a clearly inadequate screening history are those who have never been screened or have not been recently screened before age 65 years. About half of all invasive cervical cancer cases are diagnosed in women who have never been screened or have not been screened in the last 5 years (and another 10% occurs in women who did not have appropriate follow-up for an abnormal Pap smear) (23). Data from a statewide cervical cancer screening program reveal that 29% of invasive cervical cancer cases occurred in women who had never undergone cervical cytology screening (23). Data varied by age; 25% of women aged 18 to 29 years with cancer reported no previous Pap tests, and 42% of those aged 65 years or older with cancer had never been screened (24). Efforts to further reduce the burden of cervical cancer mortality can be best achieved by focusing on women who have not been adequately screened.

Although cervical cancer screening with cytology alone every 3 years is an effective strategy, HPV testing combined with cytology every 5 years is a reasonable alternative for women aged 30 to 65 years who want to potentially increase the testing interval. The USPSTF reviewed 4 fair-quality randomized, controlled trials conducted outside of the United States (NTCC [New Technology in Cervical Cancer], POBASCAM [Population Based Screening Study Amsterdam], Swedescreen, and ARTISTIC [A Randomised Trial of HPV Testing in Primary Cervical Screening]) that compared cytology alone with cytology plus HPV testing (HC2 or polymerase chain reaction) (25–30). Published data from the 4 trials show a similar number of detected cancer cases with either strategy, although differences in colposcopy referral and treatment thresholds and incomplete reporting of data from the second screening round make interpretation complex. In all 4 trials, there were slightly lower rates of CIN3+ detected in the second round of screening and fewer cancer cases in the co-testing group than in the cytology group. These differences were very small, and not all were statistically significant. In one of the largest trials with the longest follow-up (POBASCAM), more than 44 000 women in the Netherlands were randomly assigned to HPV testing with cytology or cytology alone, with repeated screening with co-testing at 5-year intervals. Cumulative data at 9 years of follow-up demonstrated a similar absolute number of cancer cases in each group (16 vs. 20 cases; $P = 0.67$) (30). Detection of CIN3+ was similar between the 2 groups. Further data on the comparability of these 2 strategies from a longitudinal cohort study by Katki and colleagues of 330 000 U.S. women were published in 2011 (31). Cumulative 5-year incidence of cervical cancer was lower in the HPV-negative and cytology-negative group than in the cytology-negative group (3.2 per 100 000 vs. 7.5 per 100 000). Detection of CIN3+ was higher in earlier screening rounds with co-testing than with cytology alone. Modeling studies support similar benefits of co-testing every 5 years and cytology every 3 years, demonstrating small differences in expected cancer cases (7.44 vs. 8.50 cases, respectively) and cancer deaths (1.35 vs. 1.55 deaths, respectively) (4).

Two large studies documenting the low risk for cytologic abnormalities after hysterectomy have been published. A cross-sectional study of more than 5000 cytology tests among women older than age 50 years documented

that identification of vaginal intraepithelial neoplasia and cancer was rare in this age group after hysterectomy (0.18 per 1000 women screened) (32). In a second study of nearly 10 000 Pap tests performed over 2 years in 6265 women who had a hysterectomy with removal of the cervix, screening yielded 104 abnormal Pap results but only 4 high-grade lesions: 3 cases of vaginal intraepithelial neoplasia and 1 case of squamous cell carcinoma of the vagina (rate of 0.42 high-grade lesion per 1000 Pap tests) (33). Whether detection of these vaginal lesions improved clinical outcomes is unknown.

Potential Harms of Screening or Treatment

Harms of screening with cytology include short-term psychological distress (anxiety, concern) related to positive results and the time and burden resulting from the evaluation of false-positive test results. Colposcopies and biopsies can occur in response to false-positive results and can be used as a proxy measure for potential downstream harms. Recent data suggest that there is a risk for adverse effects with these procedures. The results of a randomized trial comparing surveillance with immediate colposcopy among women with minimally abnormal cytology test results showed a substantially lower proportion of women in the surveillance group who reported pain (15% vs. 39%), bleeding (17% vs. 47%), or discharge (9% vs. 34%) (34).

Common treatments of high-grade precancerous lesions in the United States include cervical conization or loop electrosurgical excision, both of which can be associated with potential short- and long-term risks. Studies show that short-term risks include pain, bleeding, and discharge (2). One cohort study found that 67% of women who had loop excision reported pain, 87% reported bleeding, and 63% reported discharge (34). In addition to surgical risks inherent to excisional therapies for neoplastic lesions, treatment may increase risk for adverse outcomes of future pregnancies, including perinatal mortality, preterm delivery before 34 weeks' gestation, and low birthweight (35, 36). To date, the evidence for adverse pregnancy events after cold-knife conization or loop excision is incomplete and based largely on retrospective studies, with some inconsistencies in the categorization of the procedures performed (2).

Because the rate of positivity on HPV testing is generally higher than that for cytology, strategies that incorporate HPV testing will identify women who are HPV positive but have no evidence of a high-grade precancerous lesion; among women aged 30 years or older in primary screening settings, this proportion ranges from 4.8% to 17% (14, 16). From U.S. studies, the proportion of women undergoing co-testing who can expect to have a positive HPV test result and normal cytology findings varies by age, ranging from 11% among women aged 30 to 34 years to 2.6% among women aged 60 to 65 years (5, 6). Guidelines for management of these women have been published by ACS/ASCCP/ASCP (7). Achieving the benefits of HPV testing with cytology without increasing the risk for overtreatment will require clinicians to be responsive to currently recommended algorithms for clinical surveillance.

The lower specificity of HPV testing (that is, higher false-positive rate) raises important concerns about unnecessary diagnostic testing (that is, colposcopy) as well as identification and treatment of precancerous lesions (such as CIN2) that may regress. The POBASCAM trial reported a modestly higher cumulative detection of CIN2 with HPV testing and cytology versus cytology alone (168 vs. 127 cases) (26, 30). On the basis of these findings, 8 CIN2/CIN3 lesions would have to be treated to prevent 1 case of cervical cancer (37). Although most trials have not yet reported final cumulative colposcopy rates, data from ARTISTIC show a slightly higher proportion of colposcopy referrals in the co-testing group than with cytology alone (6% vs. 4.9%, respectively), and early results from an ongoing trial comparing screening with HPV testing versus cytology also suggests higher rates of colposcopy referrals resulting from the first round of screening with HPV (28, 29, 37). Modeling studies commissioned by the USPSTF, however, show a modest increase in colposcopy with cytology alone compared with HPV testing plus cytology and fewer overall positive test results over a lifetime of screening. Assuming screening with cytology every 3 years before age 30 years and then co-testing every 5 years in a hypothetical cohort of 1000 women, modestly fewer lifetime colposcopies could be expected with co-testing than with cytology (758 vs. 575, respectively), but more lifetime tests could be expected (approximately 5000 more lifetime tests per 1000 women with co-testing). Cumulative data from the POBASCAM trial as well as round-specific results from ongoing trials reviewed by the USPSTF should be interpreted cautiously because there are limitations in study design and diagnostic protocols (for example, cytologic thresholds for colposcopy referral and randomization schemes in subsequent rounds of screening) that could alter the balance of benefits and harms, particularly when these findings are translated to U.S. practice. Additional round-specific and cumulative data from ongoing trials may further inform the balance of potential benefits and harms of HPV testing combined with cytology versus cytology alone.

Other potential harms of HPV testing include psychological distress associated with a positive result and unnecessary evaluation of a false-positive result, as well as the time required by the patient for repeated sampling due to an inadequate or insufficient specimen. Although some women may value information about HPV status, the USPSTF found evidence of adverse short-term psychological harms associated with knowledge of HPV positivity. Four fair-quality observational studies conducted in countries with well-developed cervical cancer screening programs, including a subset of women in an Australian trial (38), examined the immediate and short-term effect of

HPV testing in more than 4000 women (2). Immediate anxiety and stress levels increased in women who tested positive for HPV compared with those who tested negative; these differences were resolved after 6 months of follow-up. Data on longer-term follow-up of women undergoing HPV testing are limited. No treatments are available to eliminate HPV. Although there is evidence of harms of strategies that incorporate HPV testing in women aged 30 to 65 years, the USPSTF concludes that there is adequate evidence that the longer screening interval for HPV testing with cytology reduces the magnitude of these harms by decreasing the opportunity for false-positive test results. Current evidence suggests that there are moderate harms of HPV testing among women younger than age 30 years. The high prevalence of HPV, higher likelihood of regression of precancerous lesions, and low incidence of cervical cancer in this age group potentiate the harms of unnecessary colposcopy and biopsy. The higher false-positive rate also increases the possibility of unnecessary treatment and the potential for adverse pregnancy outcomes.

Estimate of Magnitude of Net Benefit

The effectiveness of cervical cancer screening observed in the United States over the past several decades is attributed to the use of conventional cytology. Although there is little direct evidence from studies applicable to the U.S. population that provides an estimate of the magnitude of net benefit, observational evidence provides high certainty that the introduction of screening substantially reduces rates of cervical cancer. Recommendations regarding appropriate screening intervals seek to achieve these benefits with relatively few harms.

The harms of screening women younger than age 21 years outweigh the benefits given the high prevalence of HPV infection and associated transient cytologic abnormalities in young women; detection of these abnormalities may prompt invasive procedures and excisional treatments that have been associated with subsequent adverse pregnancy outcomes. Because of the low incidence of high-grade precancerous lesions and cervical cancer in adequately screened older women, screening for cervical cancer in women older than age 65 years is of little benefit. The harms of screening women older than age 65 years, including false-positive results and complications from follow-up and treatment of abnormalities, are judged by the USPSTF to be small. After balancing the potential benefits and harms, there is moderate certainty that screening women older than age 65 years has no benefit if they have been previously adequately screened. After hysterectomy for reasons other than a high-grade precancerous lesion or cervical cancer, screening the vagina for precancerous lesions is of little benefit and has the potential for harm due to positive test results, with subsequent invasive procedures and treatments.

Although none of the reported trials compared HPV testing with cytology-based screening as currently performed in the United States, the USPSTF was able to draw several relevant conclusions from these trials and others (26–29, 39–43), in addition to epidemiologic and natural history data (2, 18). In women younger than age 30 years, the USPSTF found evidence that the potential harms of HPV testing outweigh the potential benefits and concluded that there was no net benefit of HPV testing, alone or in combination with cytology, in this age group. This conclusion was based on the consistent and substantially higher HPV positivity rates in young women compared with older women and the potential to cause short-term adverse psychological effects and adverse pregnancy outcomes in this group of childbearing women. Detection of CIN2 is also increased with HPV testing. Many CIN2 lesions will regress, and overtreatment is a concern. In women aged 30 to 65 years, the USPSTF found that the evidence was adequate to conclude that the potential benefit of HPV testing in combination with cytology every 5 years is similar to the benefits achievable with cytology alone every 3 years.

How Does Evidence Fit With Biological Understanding?

The natural history of cervical cancer has been well-studied. Human papillomavirus infection of the cervix is generally transient, but when the infection is not cleared by an appropriate immune response and the HPV is of an oncogenic type, the infection can result in incorporation of HPV gene sequences into the host genome, which can lead to precancerous lesions. The long preclinical phase from infection to development of precancerous lesions and cervical cancer allows for the opportunity to efficiently screen for and identify precancerous lesions and treat them, thereby reducing the incidence of cervical cancer incidence and mortality.

Response to Public Comments

A draft version of this recommendation statement was posted for public comment on the USPSTF Web site from 19 October, through 30 November 2011. Many comments pointed out a lack of clarity about the harms of false-positive results and the harms of screening with cytology more frequently than every 3 years or screening women younger than age 21 years. Several comments requested clarification on how information about sexual history may affect screening. Some comments highlighted the importance of reaching women who are not being screened at all. Many comments urged the USPSTF to reconsider its draft recommendation on HPV co-testing and review new evidence that had been published since its deliberation. In response to these comments, the USPSTF clarified throughout the statement the harms that would occur from screening too frequently and in women younger than age 21 years. The USPSTF also clarified that this recommendation statement applies to women regardless of sexual history. The USPSTF agrees that the greatest effect on cervical cancer incidence and mortality would result from

efforts to screen women who have not been adequately screened, and this is stated in the Rationale and elsewhere.

After the public comment period, the USPSTF considered new evidence that was published since its initial deliberation—specifically the update of the POBASCAM results and the study by Katki and colleagues (30, 31). With this new evidence, in addition to the previously considered evidence, the USPSTF decided to recommend HPV testing combined with cytology (co-testing) as a reasonable alternative for women aged 30 to 65 years who wish to extend the screening interval beyond 3 years.

Update of Previous USPSTF Recommendation

This recommendation updates the 2003 USPSTF recommendation (44) on screening for cervical cancer. It differs from the previous recommendation in that it recommends cytology screening every 3 years among women aged 21 to 65 years. In addition, this recommendation includes more guidance on the appropriate age ranges and frequency of screening, including a new recommendation that women younger than age 21 years not be screened because the evidence shows no net benefit. The previous recommendation suggested that most of the benefit of screening could be obtained by beginning screening within 3 years of onset of sexual activity or age 21 years (whichever comes first) and screening at least every 3 years. This recommendation reaffirms the previous recommendations against screening in adequately screened women older than age 65 years and in women who have had a total hysterectomy with removal of the cervix. The current recommendation includes new evidence on the comparative test performance of liquid-based versus conventional cytology that indicates no substantial difference in test performance (that is, relative detection or absolute sensitivity or specificity) for detection of CIN2+/CIN3+. It also includes more guidance on the appropriate use of HPV testing in cervical cancer screening, including a new recommendation that women younger than aged 30 years not be screened with HPV testing. The USPSTF found new evidence that addressed the gaps identified in the previous recommendation and allowed the USPSTF to recommend HPV testing combined with cytology as an acceptable screening strategy for women aged 30 to 65 years who prefer to lengthen their screening interval beyond 3 years.

Recommendations of Others

The ACS, ASCCP, and ASCP recently published screening guidelines that are very similar to the USPSTF's recommendations (7). The ACS/ASCCP/ASCP recommend that women aged 21 to 29 years be screened with cytology (cervical cytology testing or Pap testing) alone every 3 years. Women aged 30 to 65 years should be screened with cytology and HPV testing (co-testing) every 5 years or cytology alone every 3 years. The guidelines further state that no woman should be screened every year and that women aged 21 to 29 years should not be screened with HPV testing or combined cytology and HPV testing. The American Congress of Obstetricians and Gynecologists and the American Academy of Family Physicians have previously published screening guidelines (45, 46) and are evaluating new evidence, including these recommendations on screening for cervical cancer from the USPSTF.

From U.S. Preventive Services Task Force, Rockville, Maryland.

Disclaimer: Recommendations made by the USPSTF are independent of the U.S. government. They should not be construed as an official position of the Agency for Healthcare Research and Quality or the U.S. Department of Health and Human Services.

Financial Support: The USPSTF is an independent, voluntary body. The U.S. Congress mandates that the Agency for Healthcare Research and Quality support the operations of the USPSTF.

Potential Conflicts of Interest: Disclosure forms from USPSTF members can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M12-0425.

Requests for Single Reprints: Reprints are available from the USPSTF Web site (www.uspreventiveservicestaskforce.org).

References

1. **Howlader N, Noone AM, Krapcho M, Neyman N, Aminou R, Waldron W, et al, eds.** SEER Cancer Statistics Review, 1975–2008. Bethesda, MD: National Cancer Institute; 2011. Accessed at http://seer.cancer.gov/csr/1975_2008/ on 11 October 2011.
2. **Vesco KK, Whitlock EP, Eder M, Lin J, Burda BU, Senger CA, et al.** Screening for cervical cancer: a systematic evidence review for the U.S. Preventive Services Task Force. Evidence Synthesis No. 86. AHRQ Publication No. 11-05156-EF-1. Rockville, MD: Agency for Healthcare Research and Quality; 2011.
3. **Vesco KK, Whitlock EP, Eder M, Burda BU, Senger CA, Lutz K.** Risk factors and other epidemiologic considerations for cervical cancer screening: a narrative review for the U.S. Preventive Services Task Force. Ann Intern Med. 2011;155:698-705, W216. [PMID: 22006929]
4. **Kulasingam SL, Havrilesky L, Ghebre R, Myers ER.** Screening for cervical cancer: a decision analysis for the U.S. Preventive Services Task Force. AHRQ Publication No. 11-05157-EF-1. Rockville, MD: Agency for Healthcare Research and Quality; 2011.
5. **Datta SD, Koutsky LA, Ratelle S, Unger ER, Shlay J, McClain T, et al.** Human papillomavirus infection and cervical cytology in women screened for cervical cancer in the United States, 2003-2005. Ann Intern Med. 2008;148:493-500. [PMID: 18378945]
6. **Castle PE, Fetterman B, Poitras N, Lorey T, Shaber R, Kinney W.** Five-year experience of human papillomavirus DNA and Papanicolaou test cotesting. Obstet Gynecol. 2009;113:595-600. [PMID: 19300322]
7. **Saslow D, Solomon D, Lawson HW, Killackey M, Kulasingam SL, Cain J, et al.** American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer. CA Cancer J Clin. 2012. [PMID: 22422631]
8. **Paavonen J, Naud P, Salmerón J, Wheeler CM, Chow SN, Apter D, et al; HPV PATRICIA Study Group.** Efficacy of human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by oncogenic HPV types (PATRICIA): final analysis of a double-blind, randomised study in young women. Lancet. 2009;374:301-14. [PMID: 19586656]
9. **Centers for Disease Control and Prevention, National Breast and Cervical Cancer Early Detection Program.** Screening program summaries: national aggregate. Accessed at www.cdc.gov/cancer/nbccedp/data/summaries/national_aggregate.htm on 11 October 2011.
10. **Ronco G, Cuzick J, Pierotti P, Cariaggi MP, Dalla Palma P, Naldoni C, et al.** Accuracy of liquid based versus conventional cytology: overall results of new

technologies for cervical cancer screening: randomised controlled trial. BMJ. 2007;335:28. [PMID: 17517761]
11. **Siebers AG, Klinkhamer PJ, Grefte JM, Massuger LF, Vedder JE, Beijers-Broos A, et al.** Comparison of liquid-based cytology with conventional cytology for detection of cervical cancer precursors: a randomized controlled trial. JAMA. 2009;302:1757-64. [PMID: 19861667]
12. **Coste J, Cochand-Priollet B, de Cremoux P, Le Galès C, Cartier I, Molinié V, et al; French Society of Clinical Cytology Study Group.** Cross sectional study of conventional cervical smear, monolayer cytology, and human papillomavirus DNA testing for cervical cancer screening. BMJ. 2003;326:733. [PMID: 12676841]
13. **Cárdenas-Turanzas M, Nogueras-Gonzalez GM, Scheurer ME, Adler-Storthz K, Benedet JL, Beck JR, et al.** The performance of human papillomavirus high-risk DNA testing in the screening and diagnostic settings. Cancer Epidemiol Biomarkers Prev. 2008;17:2865-71. [PMID: 18843032]
14. **Petry KU, Menton S, Menton M, van Loenen-Frosch F, de Carvalho Gomes H, Holz B, et al.** Inclusion of HPV testing in routine cervical cancer screening for women above 29 years in Germany: results for 8466 patients. Br J Cancer. 2003;88:1570-7. [PMID: 12771924]
15. **Mayrand MH, Duarte-Franco E, Rodrigues I, Walter SD, Hanley J, Ferenczy A, et al; Canadian Cervical Cancer Screening Trial Study Group.** Human papillomavirus DNA versus Papanicolaou screening tests for cervical cancer. N Engl J Med. 2007;357:1579-88. [PMID: 17942871]
16. **Kulasingam SL, Hughes JP, Kiviat NB, Mao C, Weiss NS, Kuypers JM, et al.** Evaluation of human papillomavirus testing in primary screening for cervical abnormalities: comparison of sensitivity, specificity, and frequency of referral. JAMA. 2002;288:1749-57. [PMID: 12365959]
17. **Bigras G, de Marval F.** The probability for a Pap test to be abnormal is directly proportional to HPV viral load: results from a Swiss study comparing HPV testing and liquid-based cytology to detect cervical cancer precursors in 13,842 women. Br J Cancer. 2005;93:575-81. [PMID: 16136031]
18. **Whitlock EP, Vesco KK, Eder M, Lin JS, Senger CA, Burda BU.** Liquid-based cytology and human papillomavirus testing to screen for cervical cancer: a systematic review for the U.S. Preventive Services Task Force. Ann Intern Med. 2011;155:687-97, W214-5. [PMID: 22006930]
19. **IARC Working Group on Evaluation of Cervical Cancer Screening Programmes.** Screening for squamous cervical cancer: duration of low risk after negative results of cervical cytology and its implication for screening policies. Br Med J (Clin Res Ed). 1986;293:659-64. [PMID: 3092971]
20. **Insinga RP, Glass AG, Rush BB.** Diagnoses and outcomes in cervical cancer screening: a population-based study. Am J Obstet Gynecol. 2004;191:105-13. [PMID: 15295350]
21. **Peto J, Gilham C, Deacon J, Taylor C, Evans C, Binns W, et al.** Cervical HPV infection and neoplasia in a large population-based prospective study: the Manchester cohort. Br J Cancer. 2004;91:942-53. [PMID: 15292939]
22. **Fahs MC, Mandelblatt J, Schechter C, Muller C.** Cost effectiveness of cervical cancer screening for the elderly. Ann Intern Med. 1992;117:520-7. [PMID: 1503355]
23. **Janerich DT, Hadjimichael O, Schwartz PE, Lowell DM, Meigs JW, Merino MJ, et al.** The screening histories of women with invasive cervical cancer, Connecticut. Am J Public Health. 1995;85:791-4. [PMID: 7762711]
24. **Sasieni P, Castanon A.** Call and recall cervical screening programme: screening interval and age limits. Curr Diagn Pathol. 2006;12:114-26.
25. **Ronco G, Giorgi-Rossi P, Carozzi F, Confortini M, Dalla Palma P, Del Mistro A, et al; New Technologies for Cervical Cancer screening (NTCC) Working Group.** Efficacy of human papillomavirus testing for the detection of invasive cervical cancers and cervical intraepithelial neoplasia: a randomised controlled trial. Lancet Oncol. 2010;11:249-57. [PMID: 20089449]
26. **Bulkmans NW, Berkhof J, Rozendaal L, van Kemenade FJ, Boeke AJ, Bulk S, et al.** Human papillomavirus DNA testing for the detection of cervical intraepithelial neoplasia grade 3 and cancer: 5-year follow-up of a randomised controlled implementation trial. Lancet. 2007;370:1764-72. [PMID: 17919718]
27. **Naucler P, Ryd W, Törnberg S, Strand A, Wadell G, Elfgren K, et al.** Human papillomavirus and Papanicolaou tests to screen for cervical cancer. N Engl J Med. 2007;357:1589-97. [PMID: 17942872]
28. **Kitchener HC, Almonte M, Thomson C, Wheeler P, Sargent A, Stoykova B, et al.** HPV testing in combination with liquid-based cytology in primary cervical screening (ARTISTIC): a randomised controlled trial. Lancet Oncol. 2009;10:672-82. [PMID: 19540162]
29. **Kitchener HC, Almonte M, Gilham C, Dowie R, Stoykova B, Sargent A, et al; ARTISTIC Trial Study Group.** ARTISTIC: a randomised trial of human papillomavirus (HPV) testing in primary cervical screening. Health Technol Assess. 2009;13:1-150, iii-iv. [PMID: 19891902]
30. **Rijkaart DC, Berkhof J, Rozendaal L, van Kemenade FJ, Bulkmans NW, Heideman DA, et al.** Human papillomavirus testing for the detection of high-grade cervical intraepithelial neoplasia and cancer: final results of the POBASCAM randomised controlled trial. Lancet Oncol. 2012;13:78-88. [PMID: 22177579]
31. **Katki HA, Kinney WK, Fetterman B, Lorey T, Poitras NE, Cheung L, et al.** Cervical cancer risk for women undergoing concurrent testing for human papillomavirus and cervical cytology: a population-based study in routine clinical practice. Lancet Oncol. 2011;12:663-72. [PMID: 21684207]
32. **Fox J, Remington P, Layde P, Klein G.** The effect of hysterectomy on the risk of an abnormal screening Papanicolaou test result. Am J Obstet Gynecol. 1999;180:1104-9. [PMID: 10329862]
33. **Pearce KF, Haefner HK, Sarwar SF, Nolan TE.** Cytopathological findings on vaginal Papanicolaou smears after hysterectomy for benign gynecologic disease. N Engl J Med. 1996;335:1559-62. [PMID: 8900088]
34. **Sharp L, Cotton S, Cochran C, Gray N, Little J, Neal K, et al; TOMBOLA (Trial Of Management of Borderline and Other Low-grade Abnormal smears) Group.** After-effects reported by women following colposcopy, cervical biopsies and LLETZ: results from the TOMBOLA trial. BJOG. 2009;116:1506-14. [PMID: 19583712]
35. **Kyrgiou M, Koliopoulos G, Martin-Hirsch P, Arbyn M, Prendiville W, Paraskevaidis E.** Obstetric outcomes after conservative treatment for intraepithelial or early invasive cervical lesions: systematic review and meta-analysis. Lancet. 2006;367:489-98. [PMID: 16473126]
36. **Arbyn M, Kyrgiou M, Simoens C, Raifu AO, Koliopoulos G, Martin-Hirsch P, et al.** Perinatal mortality and other severe adverse pregnancy outcomes associated with treatment of cervical intraepithelial neoplasia: meta-analysis. BMJ. 2008;337:a1284. [PMID: 18801868]
37. **Whitlock EP, Vesco KK, Lin JS, Eder M, Burda BU.** Response to human papillomavirus testing to screen for cervical cancer [Rapid Response]. Ann Intern Med. Published 3 February 2012.
38. **Kitchener HC, Fletcher I, Roberts C, Wheeler P, Almonte M, Maguire P.** The psychosocial impact of human papillomavirus testing in primary cervical screening-a study within a randomized trial. Int J Gynecol Cancer. 2008;18:743-8. [PMID: 17944916]
39. **Naucler P, Ryd W, Törnberg S, Strand A, Wadell G, Elfgren K, et al.** Efficacy of HPV DNA testing with cytology triage and/or repeat HPV DNA testing in primary cervical cancer screening. J Natl Cancer Inst. 2009;101:88-99. [PMID: 19141778]
40. **Elfgren K, Rylander E, Rådberg T, Strander B, Strand A, Paajanen K, et al; Swedescreen Study Group.** Colposcopic and histopathologic evaluation of women participating in population-based screening for human papillomavirus deoxyribonucleic acid persistence. Am J Obstet Gynecol. 2005;193:650-7. [PMID: 16150255]
41. **Kitchener HC, Almonte M, Wheeler P, Desai M, Gilham C, Bailey A, et al; ARTISTIC Trial Study Group.** HPV testing in routine cervical screening: cross sectional data from the ARTISTIC trial. Br J Cancer. 2006;95:56-61. [PMID: 16773068]
42. **Sargent A, Bailey A, Turner A, Almonte M, Gilham C, Baysson H, et al.** Optimal threshold for a positive hybrid capture 2 test for detection of human papillomavirus: data from the ARTISTIC trial. J Clin Microbiol. 2010;48:554-8. [PMID: 20007387]
43. **Sargent A, Bailey A, Almonte M, Turner A, Thomson C, Peto J, et al; ARTISTIC Study Group.** Prevalence of type-specific HPV infection by age and grade of cervical cytology: data from the ARTISTIC trial. Br J Cancer. 2008;98:1704-9. [PMID: 18392052]
44. **U.S. Preventive Services Task Force.** Screening for cervical cancer: recommendations and rationale. Rockville, MD: Agency for Healthcare Research and Quality; 2003.
45. **ACOG Committee on Practice Bulletins—Gynecology.** ACOG Practice Bulletin no. 109: Cervical cytology screening. Obstet Gynecol. 2009;114:1409-20. [PMID: 20134296]
46. **American Academy of Family Physicians.** Clinical preventive services: cervical cancer. Leawood, KS: American Academy of Family Physicians; 2003. Accessed at www.aafp.org/online/en/home/clinical/exam/cervicalcancer.html on 16 February 2012.

Appendix: U.S. Preventive Services Task Force

Members of the U.S. Preventive Services Task Force† at the time this recommendation was finalized are Virginia A. Moyer, MD, MPH, *Chair* (Baylor College of Medicine, Houston, Texas); Michael L. LeFevre, MD, MSPH, *Co-Vice Chair* (University of Missouri School of Medicine, Columbia, Missouri); Albert L. Siu, MD, MSPH, *Co-Vice Chair* (Mount Sinai School of Medicine, New York, and James J. Peters Veterans Affairs Medical Center, Bronx, New York); Kirsten Bibbins-Domingo, PhD, MD (University of California, San Francisco, San Francisco, California); Susan J. Curry, PhD (University of Iowa College of Public Health, Iowa City, Iowa); Glenn Flores, MD (University of Texas Southwestern, Dallas, Texas); Adelita Gonzales Cantu, RN, PhD (University of Texas Health Science Center, San Antonio, Texas); David C. Grossman, MD, MPH (Group Health Cooperative, Seattle, Washington); Joy Melnikow, MD, MPH (University of California, Davis, Sacramento, California); Wanda K. Nicholson, MD, MPH, MBA (University of North Carolina School of Medicine, Chapel Hill, North Carolina); Carolina Reyes, MD, MPH (Virginia Hospital Center, Arlington, Virginia); and Timothy J. Wilt, MD, MPH (University of Minnesota Department of Medicine and Minneapolis Veteran Affairs Medical Center, Minneapolis, Minnesota). Former USPSTF members who contributed to the development of this recommendation include George J. Isham, MD, MS (HealthPartners, Minneapolis, Minnesota); Rosanne M. Leipzig, MD, PhD (Mount Sinai School of Medicine, New York, New York); Bernadette Melnyk, PhD, RN (Ohio State University College of Nursing, Columbus, Ohio); George F. Sawaya, MD (University of California, San Francisco, San Francisco, California); and J. Sanford Schwartz, MD, MBA (University of Pennsylvania Medical School and the Wharton School, Philadelphia, Pennsylvania).

† For a list of current Task Force members, go to www.uspreventiveservicestaskforce.org/members.htm.

Table 1. Summary of Recommendations

Population	Page number	Recommended screening method[a]	Management of screen results	Comments
<21 Years	7	No Screening		HPV testing should NOT be used for screening or management of ASC-US in this age group.
21–29 Years	8–9	Cytology alone every 3 years	HPV-Positive ASC-US[b] or cytology of LSIL or more severe: Refer to ASCCP Guidelines [2] Cytology Negative or HPV-Negative ASC-US[b]: Rescreen with cytology in 3 years	HPV testing should NOT be used for screening in this age group
30–65 Years	9–16	HPV and Cytology ''Cotesting'' every 5 years (Preferred)	HPV-Positive ASC-US or cytology of LSIL or more severe: Refer to ASCCP Guidelines[2]	Screening by HPV testing *alone* is not recommended for most clinical settings.
			HPV Positive, Cytology Negative: Option 1 – 12-month follow-up with cotesting Option 2 – Test for HPV16 or HPV16/18 genotypes if HPV16 or HPV16/18 positive: refer to colposcopy If HPV16 or HPV16/18 negative: 12-month follow-up with cotesting	
			Cotest Negative or HPV-Negative ASC-US: Rescreen with cotesting in 5 years	
		Cytology alone every 3 years (Acceptable)	HPV-Positive ASC-US[b] or cytology of LSIL or more severe: Refer to ASCCP Guidelines [2]	
			Cytology Negative or HPV-Negative ASC-US[b]: Rescreen with cytology in 3 years	
>65 Years	16–17	No Screening following adequate negative prior screening		Women with a history of CIN2 or a more severe diagnosis should continue routine screening for at least 20 years.
After Hysterectomy	17–18	No Screening		Applies to women *without* a cervix and *without* a history of CIN2 or a more severe diagnosis in the past 20 years or cervical cancer ever.
HPV Vaccinated	18–19	Follow age-specific recommendations (same as unvaccinated women)		

[a]Women should not be screened annually at any age by any method.
[b]ASC-US cytology with secondary HPV testing for management decisions.

risk has led to the development of molecular tests for HPV* that offer increased sensitivity albeit lower specificity compared to cytology. HPV tests may better forecast which women will develop CIN3+ over the next 5–15 years than cytology [21–23]. Incorporation of HPV testing into cervical cancer screening strategies has the potential to allow both increased disease detection (improving benefits) and increased length of screening intervals (decreasing harms such as psychosocial impact of screening positive, additional clinical visits and procedures, and treatment of lesions destined to resolve). In the development of this evidence-based review and guideline update, we considered the tradeoffs of benefits and harms of screening while considering different screening modalities and ages.

*HPV refers only to high-risk HPV. Other HPV types are unrelated to cervical cancer and should not be used in cervical cancer screening. Testing for low-risk HPV types has no clinical role in cervical cancer screening or evaluation of women with abnormal cytology.

Copyright © 2012 American Society for Colposcopy and Cervical Pathology. Unauthorized reproduction of this article is prohibited.