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

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).

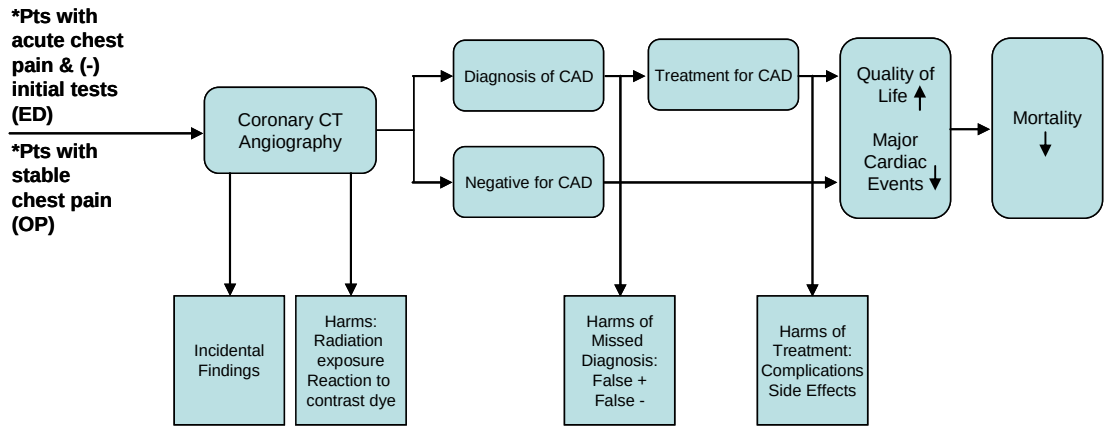
All of these alternative non-invasive diagnostic techniques measure in some way the functional impact on the heart of any underlying CAD. As noted above, none of the tests is perfect; each has the possibility of producing false positive and false negative results. Professional guidelines recognize all 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.

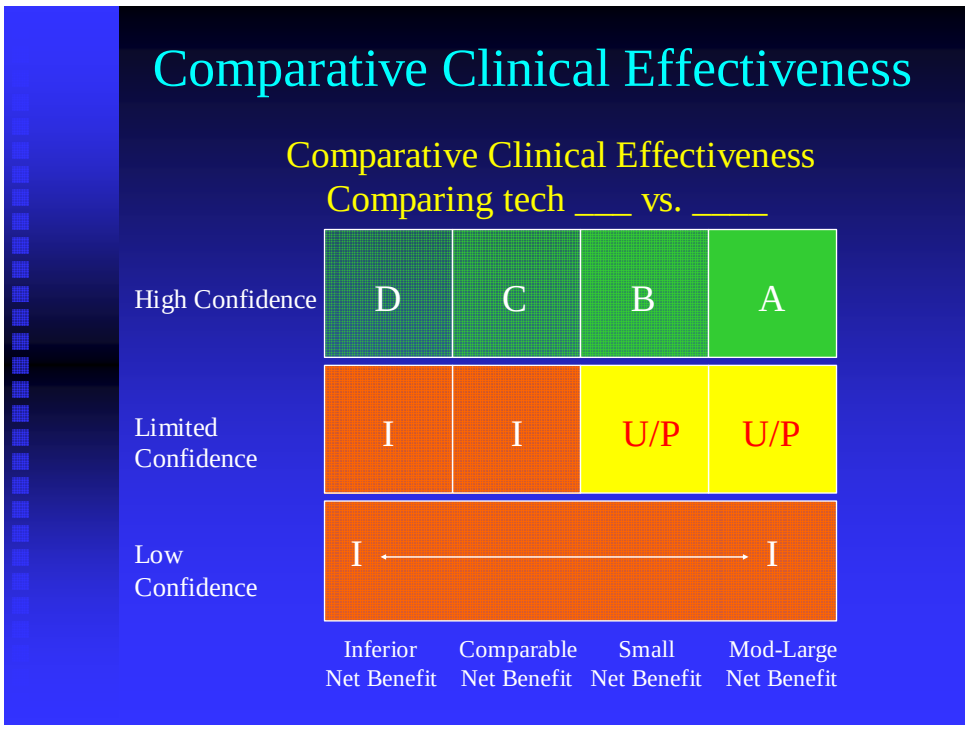
There was recognition that the evidence base for patient outcomes of CCTA in the outpatient setting was not as solid, and this was reflected in the ratings of comparative clinical effectiveness. While 4 of 11 ERG members felt that CCTA should be rated as at least "Comparable" to other non-invasive strategies, 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.

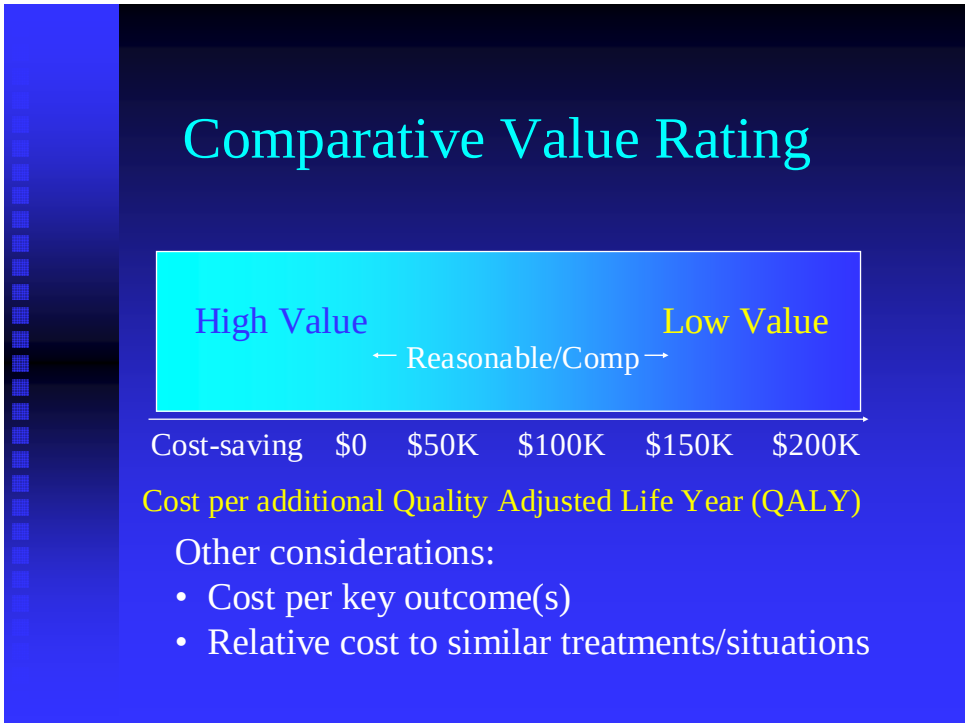
Whether judged quantitatively or qualitatively, there are two general situations that decision-making groups face in judging the balance of benefits and harms between two alternative interventions. The first situation arises when both interventions have the same types of benefits and harms. For example, 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.

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**

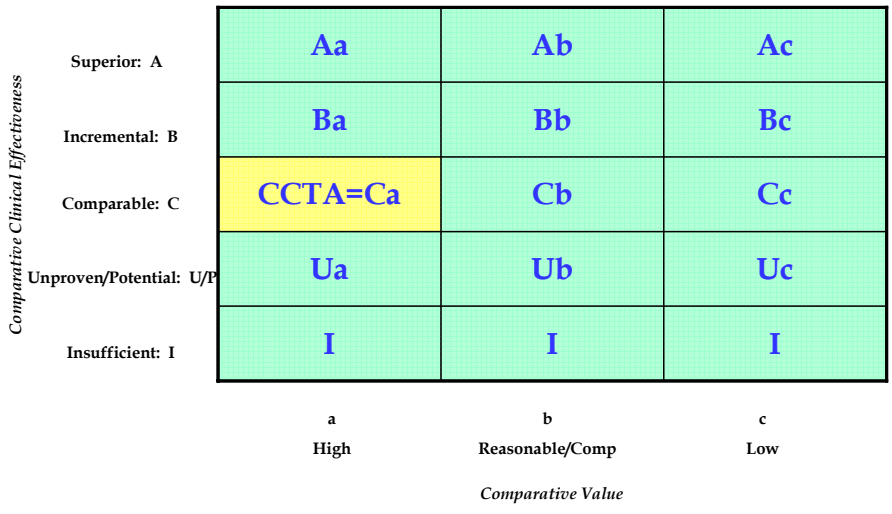
The Comparative Value of CCTA for triage of patients with acute chest pain in an ED setting is rated as:

- **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



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**

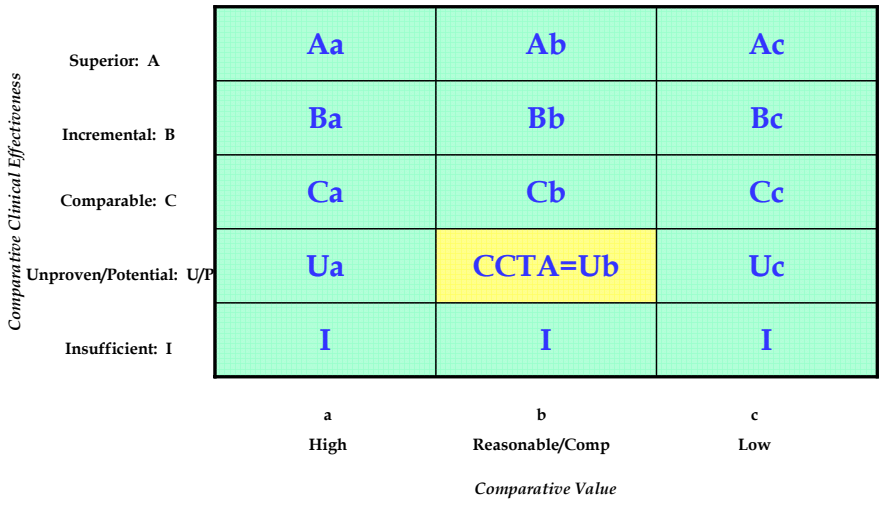
The Comparative Value of CCTA for assessment of outpatients presenting with stable chest pain is rated as:

- **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



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