Published in final edited form as:

Acad Radiol. 2012 September ; 19(9): 1049–1054. doi:10.1016/j.acra.2012.04.009.

Toward Evidence-based Decisions in Diagnostic Radiology: A Research and Rating Process for Multiple Decision-makers

Daniel A. Ollendorf, MPH, C. Craig Blackmore, MD, MPH, and Janie M. Lee, MD, MS

Institute for Clinical and Economic Review (D.A.O.) and the Institute for Technology Assessment and Department of Radiology (J.M.L.), Massachusetts General Hospital, 101 Merrimac Street, 10th Floor, Boston, MA 02114 and the Center for Health Services Research and Department of Radiology, Virginia Mason Medical Center, Seattle, WA 98101 (C.C.B.)

Abstract

The use of diagnostic imaging tests and the development of evidence-based guidelines, reviews, and other materials have both undergone substantial growth in recent years. However, the proliferation of evidence-based information has not translated into the universal deployment of medical and coverage policy for diagnostic imaging that is similarly evidence-based. One possible reason is the failure of those institutions generating evidence-based information to format findings in an accessible manner for all relevant stakeholders. The Institute for Clinical and Economic Review has developed a simple and transparent method for rating evidence that is accessible to clinicians, patients, payers, and other policy makers. The authors describe this process in relation to three imaging-based examples (computed tomographic colonography, coronary computed tomographic angiography, and positron emission tomography for dementia neuropathology). The issues raised, controversies considered, and use of the ratings in setting policy are discussed in relation to each example.

Keywords

Comparative effectiveness research; evidence-based medicine; diagnostic imaging

The past decade has featured remarkable growth in the use of the common tools and outputs of evidence-based medicine, defined as “the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients” (1). The publication of systematic reviews and meta-analyses, the preferred methodologic approach for synthesizing published evidence on the effects of medical interventions, grew by >60% between 2005 and 2009 (2). Similarly, medical societies and other purveyors of clinical practice guidelines, including the American College of Radiology, are increasingly describing their products as evidence-based (3).

During this same time period, there has also been an unprecedented increase in the use of diagnostic imaging. Annual increases in the use of computed tomography, magnetic resonance imaging, and other advanced imaging services are estimated to range from 8% to 10% in the Medicare population alone (4); Medicare expenditures for these services more than doubled between 2000 and 2006 (5). Multiple factors have been associated with this growth, including financial incentives for physician ownership of imaging devices, increased patient and clinician self-referral, and increased imaging capacity (6, 7), none of

which is well correlated with acceptance of evidence-based information to guide clinical practice.

Indeed, the submission of and payment for imaging claims appear to be largely independent of any concept of suitability for patients, despite the presence of well-accepted guidelines on appropriate imaging practice. A cross-sectional study of the American College of Radiology (ACR) Appropriateness Criteria and Medicare Part B payments for neurologic imaging found that although the most appropriate tests were twice as likely to be reimbursed as the least appropriate tests, nearly two thirds of claims for tests with “low” appropriateness for a given condition were nevertheless paid (8).

So why has the explosion in evidence-based research not resulted in widespread acceptance by decision makers to modulate imaging utilization? It is certainly possible that the variability in what is described as “evidence” is simply too broad to engender the universal trust of clinicians. A cross-sectional analysis of >300 treatment recommendations in cardiovascular management guidelines indicated that fewer than half of these recommendations were based on “high-quality” evidence (9). In addition, and in a reflection of what is found in the medical literature, many imaging guidelines focus on choices between diagnostic modalities, rather than addressing whether such testing should be performed at all. For example, the widely disseminated ACR Appropriateness Criteria provide detailed guidance on choice of imaging modality but are inconsistent on the question of whether imaging is even appropriate for a given circumstance. Finally, there is some indication that the study of gaps between evidence and practice has remained relatively constant over time, with relatively little attention paid to developing interventions to address these gaps (10).

However, it may also be the case that evidence-based data on appropriate imaging use are not universally accessible to all relevant stakeholders. For example, clinical guideline documents are rarely presented in a fashion that is digestible by patients. In addition, public and private payers may use very different approaches to estimating the potential benefit of diagnostic imaging, preferring to focus on efficiencies in clinical practice gained and improvement in long-term outcomes over the test performance statistics and other intermediate outcomes produced by most trials of imaging technologies.

The Institute for Clinical and Economic Review (ICER) was founded in 2006 for the express purpose of conducting scientifically rigorous evidence synthesis and simulation modeling, with a goal of formatting its findings in a manner that enables improved decision making by all stakeholders. ICER’s approach has been found to be useful in rating the evidence for multiple diagnostic imaging technologies; we review these examples in the following pages, describe how these ratings have been applied to medical and coverage policy, and summarize key discussions and controversies.

THE INSTITUTE FOR CLINICAL AND ECONOMIC REVIEW APPROACH

ICER’s approach to health technology assessment has been described in comprehensive detail elsewhere (11). Briefly, ICER is a hospital-based academic research group that provides independent evaluations of the comparative clinical effectiveness and comparative value of a variety of diagnostic, procedural, medical, and pharmaceutical interventions. The institute is structured to support the involvement of multiple stakeholder groups and to produce its reports in a fully transparent manner. As such, ICER’s funding comes from diverse sources, a multistakeholder advisory group is recruited to assist with each appraisal, and all of the institute’s work products are made freely available to the general public.

ICER appraisals are composed of two primary components: (1) a systematic review of the evidence on comparative clinical effectiveness, including quantitative synthesis where appropriate, and (2) a simulation model, with parameters derived in large part from the findings of the systematic review. Although the results of the modeling effort are used to evaluate the comparative economic value of the interventions under study, the models also often provide further insight into comparisons of clinical effectiveness through evaluation of downstream clinical effects, incorporation of data on patient preferences, and estimation of the joint impact of clinical benefits and potential harms.

Rating Comparative Clinical Effectiveness

ICER's process for rating comparative clinical effectiveness is illustrated in Figure 1. The underlying approach used is very similar to that used by the US Preventive Services Task Force, in that the rating is dependent on a dual judgment of (1) the level of certainty provided by the evidence and (2) the magnitude of the net health benefit provided by a given intervention in comparison with another intervention (12). All potential benefits and harms of a given intervention are considered together to make summary assessments of the net benefit of alternative interventions.

The concept of certainty is divided into three broad categories: high, moderate, and low. High certainty is typically assigned when the evidence includes consistent findings from a number of well-designed, well-conducted studies in representative populations, while the "conceptual confidence interval" around the findings in a body of evidence rated as moderate is too broad to allow firm conclusions on the direction of net benefit. Evidence rated as low is insufficient to allow any judgment on comparative clinical effectiveness.

The letter ratings shown in Figure 1 relate to summary judgments of net benefit given a particular level of certainty. With high certainty, comparative clinical effectiveness can be rated from negative (ie, an inferior net benefit vs a comparator) to superior (a moderate to substantial positive net benefit). At a level of moderate certainty, two choices are available: insufficient evidence of an inferior or comparable net benefit or unproven with potential, which indicates an emerging body of evidence with some inconsistencies, but a best estimate from that evidence indicates some level of incremental health benefit. Not surprisingly, many previous evidence reviews have focused on this latter category, as the level of residual uncertainty from the evidence is often highly correlated with the decision to prioritize the topic for review. Finally, any consideration of net health benefit for interventions with low evidence is felt to be immaterial, because this level of evidence is felt to be insufficient for making such determinations across the board.

Rating Comparative Value

As described above, the simulation models used in ICER appraisals are the primary source used to rate the comparative economic value of each intervention of interest. ICER does not use a single measure of economic impact, such as a cost-effectiveness ratio, to determine the comparative value rating. Rather, multiple measures of economic impact are considered, such as those presented in Table 1. These measures serve the dual purpose of (1) relevance to a broad set of stakeholders and (2) disaggregation of many common summary measures of economic impact to understand the key drivers of costs and effects.

For example, clinicians may be most interested in the costs incurred to prevent an adverse outcome (or produce a positive one), while patients might be most interested in the visits required and time lost from work to receive a given intervention. Within each of these measures, there may be further interest in understanding the contribution of adverse effects, follow-up and monitoring, and other issues to overall treatment costs. ICER's overarching

goal with this approach is to clarify all of the important considerations in estimating economic impact rather than have all components reported as a single summary measure with no underlying detail.

The available rating categories are high, reasonable/comparable, or low. Similar to the process for rating comparative clinical effectiveness, all economic findings, including those that address uncertainty in model outputs, are considered in making this rating. Although there are no firm parameters for the rating categories, high-value interventions are typically those that are cost saving or produce a substantial clinical benefit for marginal incremental cost, reasonable/comparable interventions produce either a smaller incremental clinical benefit or a large benefit but at a relatively high cost, and low-value interventions produce relatively little clinical benefit at a substantial cost.

Integrated Evidence Rating

After the independent evaluation of comparative clinical effectiveness and comparative value, these assessments are then brought together to form an integrated evidence rating (Fig 2). The logic for the integrated rating follows that of the ratings for the effectiveness and value ratings, with one exception. Interventions for which evidence on comparative clinical effectiveness is considered to be insufficient do not typically have any comparative value rating assigned, because there is nothing to gain by evaluating the potential economic impact if the true clinical benefit is unknown. Exceptions have been made, however, particularly for interventions already in widespread use despite the lack of evidence.

CASE STUDIES USING THE INSTITUTE FOR CLINICAL AND ECONOMIC REVIEW RATING APPROACH

Three case studies illustrating the application of the ICER rating approach to imaging policy are described below. As will be made clear, discussions centered not only on the applicability of ICER ratings to each situation but also on the specific concerns with the evolutionary stage of each body of evidence.

Computed Tomographic Colonography (CTC)

One of ICER's first appraisals compared computed tomographic colonography (CTC) with optical colonoscopy for detection of polyps >5 mm in size (13). ICER appraised CTC independently as part of its own self-contained process and subsequently as a topic for the State of Washington's Health Care Authority (HCA), which uses findings from commissioned health technology assessments to inform coverage decisions for the approximately 1 million Washington citizens for whom the state purchases health insurance (14).

The CTC topic highlighted many challenges frequently encountered when reviewing evidence related to an emerging imaging-based diagnostic test. For one, diagnostic accuracy is typically assessed on a "one-time" basis in comparison with a gold standard, without any information on the appropriate screening interval for future testing (15). In addition, emerging evidence is usually populated by relatively small studies of highly selected patients conducted at academic centers by those most experienced in the use of the device. Such studies are at high risk for spectrum bias, because included patients do not represent the full spectrum of risk for the condition of interest that would be expected in the target population for screening. These phenomena raise questions about the generalizability of findings when applied to broader populations of screeners and the screened (16, 17).

Deliberations during both the ICER appraisal and Washington HCA process focused on the comparability of CTC's diagnostic accuracy to that of optical colonoscopy, the relative safety of the two approaches, and the cost of each strategy. Because no studies have been conducted directly comparing outcomes of these two modalities, interpretations of effectiveness were limited to comparisons of pooled estimates of diagnostic accuracy as well as findings from ICER's modeling effort. The evidence on CTC and optical colonoscopy was felt to suggest comparable accuracy overall, and even slightly better accuracy for CTC if the screening interval was reduced from 10 to 5 years (15). The discussion of safety suggested that there were trade-offs to consider for each approach (ie, radiation risk and incidental findings for CTC vs bowel perforation for colonoscopy) but that both procedures appeared to be reasonably and comparably safe.

For all of the above reasons, CTC's comparative clinical effectiveness was considered to be comparable with that of optical colonoscopy (13, 18). Discussions regarding cost were more complex and illustrate the benefits of including a model-based analysis in the appraisal. First, it was recognized that positive findings on CTC would require a second procedure for polyp removal. Additionally, with effectiveness and safety deemed to be comparable, CTC would appear to be cost-effective or cost-saving only at prices half as low as or lower than those of optical colonoscopy, given the reduced screening interval for CTC. The ICER appraisal allowed for different "value" ratings on the basis of varying assumed prices for CTC.

Impact of appraisal—The Washington HCA subsequently issued a noncoverage decision for CTC on the basis of its own additional reimbursement data suggesting that CTC claims were reimbursed at a rate approximately 25% higher than that for optical colonoscopy (18). In this particular case, it is important to note that optical colonoscopy is covered by Medicare and therefore has a mechanism in place for standardized reimbursement. CTC, on the other hand, is not yet a covered service, so reimbursement by definition is both more variable and based on "usual and customary" fees. The higher fees paid by the state for CTC are therefore in large part an artifact of differences in reimbursement methods, rather than an observed difference in the resource intensity required for the procedure.

Coronary Computed Tomographic Angiography (CCTA)

As with CTC, ICER's appraisal of CCTA for coronary artery disease was also conducted as a stand-alone project (19) and presented to the Washington HCA (20). Consultation with clinical experts and payers suggested that use of CCTA as a diagnostic tool should be considered in two distinct settings: (1) triage of acute chest pain of unknown origin in the emergency department (ED) and (2) outpatient investigation of stable chest pain symptoms. In both settings, the target population included patients at low to intermediate risk for coronary artery disease (10%–30%), because those at higher levels of risk were assumed to be candidates to proceed directly to invasive coronary angiography.

In addition to general concerns regarding emerging evidence for imaging tests described previously, CCTA also raised the issue of rapid and seemingly constant technological evolution. On the basis of recommendations from its advisory group, ICER decided to limit the review to studies using scanners with 64-slice or greater precision, because this represented a generally accepted "cut point" that made comparisons of CCTA with other noninvasive test strategies feasible. However, it was also recognized that there remained substantial variation in test protocol, scanner performance by vendor, and procedures to limit radiation exposure across centers and studies and that these elements appeared to be evolving rapidly.

Review of the evidence on CCTA indicated that study results differed by setting. Although the majority of available studies were performed in an outpatient setting, the results of ED studies were found to be more compelling. Studies that attempted to compare the impact of CCTA with that of other noninvasive diagnostic strategies on patient outcomes were conducted in the ED only; these included a randomized controlled trial comparing CCTA with serial enzyme and nuclear stress testing (21). Although findings were mixed on whether CCTA would reduce the numbers of invasive angiographies required for a given population, CCTA's high negative predictive value allowed faster ED discharge compared with standard triage care, without any major increase in adverse effects or the rate of major cardiovascular events (22).

In contrast, nearly all of the 44 outpatient studies available focused on diagnostic accuracy alone, including three large multicenter studies (23–25). Diagnostic findings suggested uniformly high sensitivity, but inconsistent specificity, which may have been influenced by variability in disease risk, patient age, interreader reliability, and other factors.

The disparate nature of the evidence by setting led to two different judgments using the ICER rating matrix. Use of CCTA in the ED was found to be comparable with standard triage care, given equivalent or superior diagnostic performance and multiple unknowns (eg, reductions in need for invasive angiography, radiation risks). In the outpatient setting, however, the lack of outcome data beyond diagnostic accuracy provided a lower level of evidence for clinical efficacy. In addition, expected reductions in diagnostic accuracy in those at lower levels of coronary artery disease risk reduced the certainty around CCTA's effectiveness in comparison with other forms of noninvasive testing, leading to a rating of unproven with potential.

Data from the modeling effort supported the clinical assessment. In the ED, CCTA was found to be a high-value option, because its potential to speed discharge in many patients would offset the additional cost of the test and potentially result in cost savings to the hospital and payer. In the outpatient setting, however, CCTA's performance was found to be comparable with that of other noninvasive strategies, but at a somewhat higher cost. As a result, a value of reasonable/comparable was assigned.

Impact of appraisal—The Washington HCA subsequently concluded that the only clear evidence in favor of CCTA was for investigation of acute chest pain and limited coverage to investigation of acute chest pain in an ED or hospital setting (20).

Positron Emission Tomography for Dementia Neuropathology

The third and most recent example of the application of ICER's processes to diagnostic imaging relates to an ongoing appraisal of the potential use of positron emission tomographic scans to diagnosis Alzheimer's disease in both symptomatic and asymptomatic individuals. Recent advancements in both positron emission tomography-based and other biomarker tests have generated an immense amount of interest from clinicians, patients, and other interested parties, as they have the potential to identify persons at high risk for developing Alzheimer's disease years, and in some cases decades, before symptoms develop (26).

Although these developments have energized a community clearly eager for any breakthrough in the management of a serious and heart-wrenching illness, there are at least as many questions as answers regarding the current value of biomarker testing. Biostatistician Margaret Pepe (27) outlined six simple criteria for a useful diagnostic or screening test: (1) the disease should be serious or potentially so, (2) the disease should be relatively prevalent in the target population, (3) the disease should be treatable, (4) treatment

should be available to those who test positive, (5) the test should not harm the individual, and (6) the test should accurately classify diseased and nondiseased individuals. For screening tests in particular, a seventh criterion should be added: the effectiveness of treatment should be dependent on early diagnosis (28). Otherwise, treatment could be reserved until symptoms become manifest, and screening could be avoided.

Even if one considers these criteria as existing on a continuum (ie, not all criteria would need to be completely satisfied for the test to be considered useful), there are challenges with universal acceptance of Alzheimer's diagnostic testing in the current environment, regardless of early results on test accuracy. First, there are no currently available treatments with any evidence of sustained effectiveness in modifying disease progression or even addressing symptoms (29, 30), so criteria 3 and 4 are immediately called into question. Also, lacking is evidence that early treatment with available options is more effective than treatment in patients with more pronounced symptoms (criterion 7). In addition, Although these biomarker tests might be considered minimally invasive or noninvasive (except perhaps testing of cerebrospinal fluid via lumbar puncture), they are not without potential harm (criterion 5). For example, what are the potential radiation effects of positron emission tomographic scanning? How important are these effects when considering one-time testing for diagnosis in a symptomatic 75-year-old patient versus screening asymptomatic individuals every 10 years beginning at age 50?

Finally, considerations of test accuracy must take into account disease prevalence as well as the relative value of positive versus negative diagnosis. For example, if testing is conducted only among the elderly, among whom disease prevalence is on the order of 10%, a highly accurate test (99% sensitive and specific) would yield robust positive and negative predictive values. However, screening conducted on a general population basis (with prevalence closer to 2%) using the same test, although effectively ruling out Alzheimer's for virtually all patients who test negative, would correctly diagnose Alzheimer's in only two thirds of those who test positive (31). A slight reduction of test performance to 95% sensitivity and specificity would yield a positive predictive value of only 28% (31).

The level of uncertainty regarding the benefits and potential harms of Alzheimer's diagnostic testing in the current environment, the appropriate target populations for testing, and the potential effects of test results on clinician, patient, and caregiver decision making is such that the ICER appraisal process must operate very differently. Rather than measurement of the long-term clinical and economic impacts of various test-and-treat strategies, the chief intent of modeling in this case will be to highlight major gaps in our understanding of the impact of testing on clinical decisions and identify measures and studies to address these. For example, in the absence of effective treatment, attention may turn to the so-called psychic value of testing: the ability to cognitively adjust to an eventual diagnosis and conduct advance planning or the reassurance that comes with a negative result. However, there is little guidance from the literature on how best to measure such value, how frequently to measure it, and how long the value is sustained (32). The appraisal's purpose in this case may be to identify appropriate instruments and study designs to measure psychic value, the types of individuals in whom such value should be measured, and other related concerns. Similarly, the systematic review on test accuracy may yield important information on variation in testing procedures, diagnostic thresholds, and interpretation of findings, all of which could be addressed in a set of recommendations to standardize future study.

CONCLUSIONS

Although the conduct of comparative effectiveness research studies has increased substantially in recent years, this growth has not been accompanied by a complementary initiative to create an accessible process for rating evidence and using these ratings to make informed decisions. This is of particular concern in diagnostic radiology, which has also seen increased utilization accompanied by growing concerns regarding appropriateness.

The Institute for Clinical and Economic Review was created to address such concerns using a transparent, multistakeholder process, and featuring a straightforward evidence rating matrix intended to be accessible to patients, clinicians, payers, and policy makers alike. Experience thus far has shown that the rating matrix and process facilitate clinical decision making, particularly with diagnostic technologies already in widespread or expanding use. The utility of this approach for truly emerging technologies such as dementia diagnostics is not yet known but appears to have promise as a method for identifying the type of evidence that will reduce uncertainty and enable decision making in the future.

Acknowledgments

Dr Lee's participation was supported in part by grant K07 CA128816 from the National Institutes of Health (Bethesda, MD).

References

1. Sackett D, Rosenberg W, Gray J, et al. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996; 312:71–72. [PubMed: 8555924]
2. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med*. 2010; 7:e1000326. [PubMed: 20877712]
3. American College of Radiology. ACR Appropriateness Criteria®. Available at: <http://www.acr.org/ac>. Accessed June 3, 2012.
4. Bhargavan M, Sunshine JH. Utilization of radiology services in the United States: Levels and trends in modalities, regions, and populations. *Radiology*. 2005; 234:824–832. [PubMed: 15681686]
5. US Accountability Office. Medicare: trends in fees, utilization, and expenditures for imaging services before and after implementation of the Deficit Reduction Act of 2005. Washington, DC: US Government Accountability Office; 2008.
6. Mongan JJ, Ferris TG, Lee TH. Options for slowing the growth of health care costs. *N Engl J Med*. 2008; 358:1509–1514. [PubMed: 18385503]
7. Levin DC, Rao VM, Kaye A. Turf wars in radiology: recent actions against self-referral by state governments, commercial payers, and Medicare—hope is on the horizon. *J Am Coll Radiol*. 2008; 5:972–977. [PubMed: 18755437]
8. Siström CL, McKay NL. Evidence-based imaging guidelines and Medicare payment policy. *Health Serv Res*. 2008; 43:1006–1024. [PubMed: 18454778]
9. McAlister FA, van Diepen S, Padwal RS, et al. How evidence-based are the recommendations in evidence-based guidelines? *PLoS Med*. 2007; 4:e250. [PubMed: 17683197]
10. Evensen AE, Sanson-Fisher R, D'Este C, et al. Trends in publications regarding evidence-practice gaps: a literature review. *Implement Sci*. 2010; 5:11. [PubMed: 20181079]
11. Ollendorf DA, Pearson SD. An integrated evidence rating to frame comparative effectiveness assessments for decision makers. *Med Care*. 2010; 48(suppl):S145–S152. [PubMed: 20473206]
12. Sawaya GF, Guirguis-Blake J, LeFebvre M, et al. Update on the methods of the US Preventive Services Task Force. *Ann Intern Med*. 2007; 147:871–875. [PubMed: 18087058]
13. Scherer, R.; Knudsen, A.; Pearson, SD. Final appraisal document: CT colonography for colorectal cancer screening. Available at: <http://www.icer-review.org/index.php/Reports/ctc.html>. Accessed June 3, 2012

14. Blackmore CC, Budenholzer B. Applying evidence-based imaging to policy: the Washington state experience. *J Am Coll Radiol*. 2009; 6:366–371. [PubMed: 19394578]
15. Pearson SD, Knudsen A, Scherer RW, et al. Assessing the comparative effectiveness of a diagnostic technology: CT colonography. *Health Aff (Millwood)*. 2008; 27:1503–1514. [PubMed: 18997205]
16. Nicholson FB, Barro JL, Bartram CI, et al. The role of CT colonography in colorectal cancer screening. *Am J Gastroenterol*. 2005; 100:2315–2323. [PubMed: 16181386]
17. Deeks JJ. Systematic reviews of evaluations of diagnostic and screening tests. *BMJ*. 2001; 323:157–162. [PubMed: 11463691]
18. Washington State Health Care Authority. Health Technology Clinical Committee: findings and coverage decision. Available at: http://www.hta.hca.wa.gov/documents/decision_finding_vc_final_090308.pdf. Accessed June 3, 2012.
19. Ollendorf, DA.; Göhler, A.; Pearson, SD., et al. Final appraisal document: coronary computed tomographic angiography for detection of coronary artery disease. Available at: <http://www.icer-review.org/index.php/Reports/ccta.html>. Accessed June 3, 2012.
20. Washington State Health Care Authority. Health Technology Clinical Committee: findings and coverage decision. Available at: http://www.hta.hca.wa.gov/documents/findings_decision_ccta_051209.pdf. Accessed June 3, 2012.
21. 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. *J Am Coll Cardiol*. 2007; 49:863–871. [PubMed: 17320744]
22. Ollendorf DA, Kuba M, Pearson SD. The diagnostic performance of multi-slice coronary computed tomographic angiography: a systematic review. *J Gen Intern Med*. 2010; 26:307–316. [PubMed: 21063800]
23. 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; 52:1724–1732. [PubMed: 19007693]
24. Miller JM, Rochitte CE, Dewey M, et al. Diagnostic performance of coronary angiography by 64-row CT. *N Engl J Med*. 2008; 359:2324–2336. [PubMed: 19038879]
25. Meijboom WB, Meijs MF, Schuijf JD, et al. Diagnostic accuracy of 64-slice computed tomography coronary angiography. *J Am Coll Cardiol*. 2008; 52:2135–2144. [PubMed: 19095130]
26. Freundlich, N. Diagnosis without treatment: the perils of new tests for early Alzheimer's disease. Available at: <http://www.healthbeatblog.com/2010/07/diagnosis-without-treatment-the-perils-of-new-tests-for-early-alzheimers-disease.html>. Accessed December 22, 2011.
27. Pepe, MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York: Oxford University Press; 2003.
28. Obuchowski NA, Graham RJ, Baker ME, et al. Ten criteria for effective screening: their application to multislice CT screening for pulmonary and colorectal cancers. *AJR Am J Roentgenol*. 2001; 176:1357–1362. [PubMed: 11373191]
29. Birks J. Cholinesterase inhibitors for Alzheimer's disease (review). *Cochrane Database Syst Rev*. 2006; 25:CD005593. [PubMed: 16437532]
30. Loy C, Schneider L. Galantamine for Alzheimer's disease and mild cognitive impairment (review). *Cochrane Database Syst Rev*. 2006; 25:CD001747. [PubMed: 16437436]
31. Mattson N, Zetterberg H. Future screening for incipient Alzheimer's disease—the influence of prevalence on test performance. *Eur Neurol*. 2009; 62:200–203. [PubMed: 19602894]
32. Lee DW, Neumann PJ, Rizzo JA. Understanding the medical and nonmedical value of diagnostic testing. *Value Health*. 2010; 13:310–314. [PubMed: 19744295]

Figure 1.

Evidence-based medicine matrix for rating evidence on comparative clinical effectiveness. A = superior: high certainty of a moderate to large net health benefit; B = incremental: high certainty of a small net health benefit; C = comparable: high certainty of a comparable net health benefit; D = inferior: high certainty of an inferior net health benefit; U/P = unproven with potential: moderate certainty of a small or moderate to large net health benefit (this category is meant to reflect technologies whose evidence provides high certainty of at least comparable net health benefit or moderate certainty suggesting a small or moderate to large net health benefit); I = insufficient: the evidence does not provide high certainty that the net health benefit of the technology is at least comparable to that provided by the comparator(s).

Comparative Clinical Effectiveness	Superior: A	Aa	Ab	Ac
	Incremental: B	Ba	Bb	Bc
	Comparable: C	Ca	Cb	Cc
	Inferior: D	Da	Db	Dc
	Unproven/Potential: U/P	Ua	Ub	Uc
	Insufficient: I	I	I	I
		a	b	c
		High	Reasonable/Comp	Low
		Comparative Value		

Figure 2.
Institute for Clinical and Economic Review integrated evidence rating matrix.

TABLE 1

Measures of Economic Impact Considered in Appraisals Conducted by the Institute for Clinical and Economic Review

Measure	Of Primary Interest to
Resource use (eg, test frequency, need for hospitalization)	Patients, clinicians
Cost to achieve treatment success	Clinicians
Cost to prevent adverse outcome	Clinicians
Cost per life-year gained	Policy makers, payers
Cost per quality-adjusted life-year gained	Policy makers, payers
Budgetary impact	Payers, health systems
Manpower trade-offs	Health systems