Cost-effectiveness Analysis in the Assessment of Diagnostic Imaging Technologies

In many ways, diagnostic technologies differ from therapeutic medical technologies. Perhaps most important, diagnostic technologies do not generally directly affect long-term patient outcomes. Instead, the results of diagnostic tests can influence the care of patients; in that way, diagnostic tests may affect long-term outcomes. Because of this, the benefits associated with the use of a specific diagnostic technology will depend on the performance characteristics (eg, sensitivity and specificity) of the test, as well as other factors, such as prevalence of disease and effectiveness of available treatments for the disease in question. The fact that diagnostic tests affect short-term, or “surrogate,” outcomes, rather than long-term patient outcomes makes evaluation of these tests more complicated than the evaluation of therapeutic technologies. This article will trace the history of technology assessment in medicine, address the role of cost-effectiveness and decision analysis in health technology assessment, and describe unique features and approaches to assessing diagnostic technologies. The article will then conclude with a consideration of the limits of medical technology assessment.

The explosive growth in medical technology and procedures (1–3) during the past several decades has provided physicians with an unparalleled ability to diagnose abnormalities and treat patients. In many cases, however, new technologies and procedures have diffused widely without careful assessment of their appropriate roles in the care of patients. The uncertainty around the effectiveness of a diagnostic technology can result in controversy; examples of this include the use of computed tomography (CT) in lung cancer screening and mammography in women aged 40–49 years (4–12). The increase in available medical technologies may also have contributed to the nearly continuous increase in health care spending since the 1960s (13), the burden of which has fallen increasingly on the public sector. For instance, 45.9% of national health expenditures in the United States in 2002 were from federal or state funds (14). Heightened awareness of these issues has focused attention on constraining the growth of health care costs. Thus, as both our needs and our capabilities continue to grow in the face of constrained resources, we are increasingly faced with the dilemma of choosing either the best tests and procedures or the most cost-effective tests and procedures from among worthy alternatives, rather than merely determining which tests are effective and which are not. In particular, we are confronted with the task of identifying population subgroups for which a diagnostic procedure is clinically beneficial and cost-effective, so that clinical algorithms can be used to consider individual patient characteristics.

As we progress further into the current “era of assessment and accountability” (15), more people agree that we must improve and expand our technology assessment techniques and efforts (16–20). Multiple journals are now devoted to technology assessment; however, neither technology assessment nor its specific application to health care is a new activity. Physicians and others who provide or pay for health care have long been engaged in informal technology assessment with attempts to better understand the effects and relative merits of medical interventions. Clinical research, however, has improved considerably, and we now have tools to gather data and compare the efficacies of new and existing technologies. For many of these assessments, the randomized controlled trial (RCT) is the tool of choice.

RCTs, described as “the crown jewel of traditional technology assessment” (21), have
thus, it may be impossible to draw conclusions about narrowly defined groups of patients and include all conditions for which the procedures might be performed. Furthermore, randomized trials are not particularly well suited for use in the evaluation of diagnostic imaging tests. Withholding a noninvasive imaging test that might provide useful diagnostic information creates difficult ethical dilemmas. RCTs also cannot be used to examine the myriad of possible combinations of tests or threshold values that might be performed in a given clinical situation, nor can they be used to predict the usefulness of a test, should treatments for the condition improve.

Diagnostic technologies differ in many ways from therapeutic medical technologies. Most important, diagnostic technologies do not generally directly affect long-term patient outcomes. Instead, the results of diagnostic tests can influence the care of patients, and in that way, they may affect long-term outcomes. Because of this, the benefits associated with the use of a specific diagnostic technology will depend on the performance characteristics (eg, sensitivity and specificity) of the test and on other factors, such as the prevalence of disease and the effectiveness of available treatments for the disease in question. The fact that diagnostic tests affect short-term, or “surrogate,” outcomes rather than long-term patient outcomes makes the evaluation of these tests more complicated than the evaluation of therapeutic technologies.

This article will trace the history of technology assessment in medicine, address the role of cost-effectiveness and decision analysis in health technology assessment, and describe unique features and approaches in the assessment of diagnostic technologies. We will then conclude with a consideration of the limits of medical technology assessment.

**BRIEF REVIEW OF AMERICAN MEDICAL TECHNOLOGY ASSESSMENT SINCE 1960**

The phrase “technology assessment” first entered into public policy debates during the 1960s and 1970s, principally regarding issues such as nuclear waste disposal and efforts to reduce the health risks from exposure to environmental toxins. During this period, studies by the National Academy of Sciences (23), National Academy of Engineering (24), and National Academy of Public Administration (25) created the foundation on which Congress established the Office of Technology Assessment in 1972 (26). This was followed in 1974 by the establishment of a Health Program within the Office of Technology Assessment; soon thereafter, the growth of technology assessment in medicine began to escalate.

In the 1970s, technology assessment could include evaluations of any or all of the following end points: feasibility, safety, efficacy, appropriateness, cost, or cost-effectiveness. Technology assessment in medicine, however, has focused primarily on technical and clinical end points, such as feasibility, safety, and efficacy, despite a growing recognition of the need to broaden the scope of the investigations. Additionally, and in somewhat of a contrast to the technology assessment activities in other fields (eg, environmental health), medical technology assessment was focused more on primary rather than secondary effects of the technology being considered (27).

Throughout the 1970s and 1980s, the continued development of expensive medical technologies, such as computed tomography (CT), may have provided a powerful stimulus for broadened technology assessment in medicine (28–35). These new technologies offered substantial improvements in diagnosis, treatment, or both, when compared with the preceding technology, but the improvements often came with an expensive price tag. Several other developments also helped to catalyze technology assessment as a major component of medical research. Prominent among these was the creation by Congress of the National Center for Health Care Technology (NCHCT) in 1978 (36). The NCHCT was intended to become the principal federal technology assessment agency (27). It was located, along with the National Center for Health Services Research and the National Center for Health Statistics, in the Office of the Assistant Secretary of Health, and the NCHCT was responsible for analyzing emerging and existing medical technologies. The NCHCT was also responsible for advising the Health Care Financing Administration on issues related to Medicare coverage.

In many ways, the origins of a concerted technology assessment research enterprise in this country can be traced to the NCHCT (27). Powerful organizations, such as the American Medical Association and the Health Industry Manufacturers Association, expressed opposition to its formation, however, and they lobbied against its continued support. Ultimately, the combination of their opposition and a changing political landscape...
prevailed. Funding for the NCHCT was terminated in 1981, and the NCHCT ceased to exist shortly thereafter.

Though the demise of the NCHCT was a setback for federally supported technology assessment research in medicine, it did not signal the end. In 1984, the National Center for Health Services Research was expanded and renamed the National Center for Health Services Research and Health Care Technology Assessment, and in 1986, the Institute of Medicine of the National Academy of Sciences established the Council on Health Care Technology. During this period, the scope of medical technology assessment continued to expand, and end points relating to quality of care, patient well-being, functional health status, and costs were featured more prominently in the national research agenda (19,20,27,37–42). As a result, assessments of diagnostic imaging technologies, and specifically their effects on patient management and outcomes, began to appear with increasing frequency (30–35,37,38,43).

The government was not alone in supporting technology assessment in medicine. Private sector involvement also began to expand. For example, Blue Cross and Blue Shield established its own technology evaluation program, which developed specific criteria for use in the assessment of medical technologies (44). Notably, Blue Cross and Blue Shield required evidence that a net positive effect on health outcomes could be attained in general practice, which often differs from clinical trial settings. The American Medical Association, despite—or perhaps because of—its prior opposition to the NCHCT, also entered the fray, establishing its own Diagnostic and Therapeutic Technology Assessment program (45).

Medical technology assessment received its next major boost with the Omnibus Budget Reconciliation Act of 1989, which resulted in a substantial increase in funding to support medical technology assessment and paved the way for the establishment of the Agency for Health Care Policy and Research (now the Agency for Health Care Research and Quality). This agency subsumed the activities and responsibilities of the National Center for Health Services Research. It was specifically charged with developing research databases, facilitating the dissemination of research results, developing practice guidelines, and evaluating medical practice in general. The Agency for Health Care Policy and Research (now the Agency for Health Care Research and Quality) continued to play an important role in medical technology assessment throughout the 1990s and into the 21st century.

The establishment of such high-profile agencies has spawned a host of smaller agencies and organizations and has served to increase support for broadening the scope of medical research. As described previously, medical technology assessment can include a global assessment of the effect of medical practice on patient well-being (19,21) by including issues such as functional well-being (46–48), quality of life (49,50), and patient preferences (51,52).

As medical researchers began to broaden the scope of their assessments, they realized that methods necessary for expanded technology assessments did not yet exist, were not fully developed, or belonged to an entirely different group of scientists than those who had traditionally assessed the safety and efficacy of medical interventions. New skills and approaches were needed, and new scientists had to be recruited to participate in the studies. Not only did the new technology assessments require physicians and biostatisticians, they also required epidemiologists, economists, quality of life experts, decision scientists, and health policy experts. Other challenges also had to be faced. The new technology assessments required access to more and different data sets than previous studies needed. Along with the increased demand for data came the need for improved methods to process, analyze, and present these data in an accessible format. Along with the desire to better understand the trade off between cost and efficacy that is faced when choosing between competing technologies (in the situation of restricted resources) came the need to develop and implement formal methods of cost-effectiveness analysis (CEA) appropriate for evaluation of health care technologies.

To better understand the state of medical technology assessment and to promote further methodologic development in the assessment of the economic value of health technologies, the U.S. Public Health Service in 1993 convened the Panel on Cost-Effectiveness in Health and Medicine. The formation of this panel was motivated by a recognition of the importance of CEA and technology assessment in health and medicine and a concern for the tremendous variations in the methods and quality of currently available analyses. This group of 13 scientists and scholars (none of whom were government employees) with expertise in the techniques and applications of CEA was charged with assessing the current state of the science and making consensus-based recommendations that might improve methods and promote the transparency of research and comparability of results. The group met 11 times over a period of 2½ years and produced a book and several articles dealing with critical issues in CEA (53–57). The focus of the panel was on policy decisions and resource allocation at a broad level, with particular attention paid to reaching consensus regarding the conduct of CEAs and proposing steps that should be taken to address the remaining unresolved issues. Principal among the panel’s recommendations was the importance of including a reference case analysis in all CEAs. The reference case analysis permits results of different CEAs to be compared and allows decision makers to better understand the implications of their resource allocation decisions when funding is directed to one program or technology rather than another.

**CEA in Medical Technology Assessment**

Medical care now consumes more than 14% of the gross domestic product in the United States, which is more than any other nation in the world (14); by 2012, it could reach 17.7% (58). Furthermore, the rate at which spending increases each year has begun to rise recently after a temporary pause (13,14). Despite the ever-expanding array of expensive new technologies, it is clear that the resources allocated to health care can no longer be allowed to expand without limit. Heightened awareness of these issues has resulted in increased pressure on medical researchers and policy makers to develop and apply new analytic methods. The hope has been that if we could only understand the trade off between clinical benefits and costs, we might be able to make more informed decisions and rein in the spiraling health care costs.

CEA is a method used to evaluate health outcomes and costs of different medical technologies and procedures relative to one another (53,55,59). It is a tool that analysts and decision makers can use to compare competing options and select those that best meet their needs within budget constraints. CEA is used to evaluate technologies and procedures of interest through the use of the (incremental) cost-effectiveness ratio. In this ratio, changes in resource use (“costs”) relative to a relevant alternative are summed in the numerator, while changes
in health effects are summed in the denominator. Thus, the denominator includes productivity costs of morbidity and functional limitation. The most commonly used measures of health outcome are measures of survival, such as the number of lives or life-years saved; however, quality of life may be incorporated into the analysis by using measures such as quality-adjusted life-years (49,59–63) or healthy-year equivalents (50). In addition, adjustments may be made for the timing of future benefits or costs by discounting these in an appropriate manner (64–67).

When cost-effectiveness analytic techniques are used, it is possible to compare different health interventions, taking into account both cost and effectiveness (68). The resulting incremental cost-effectiveness ratios indicate the cost of each additional unit of health outcome (eg, quality-adjusted life-years) that one might wish to “purchase” by investing in health care interventions that are more expensive, more effective, or both. Strategies with lower cost-effectiveness ratios are considered to be more cost-effective than strategies with higher cost-effectiveness ratios.

Analysis generally proceeds in the following manner: First, we determine the best estimates of the parameters of interest for the analysis to perform so-called base-case analysis. Then, we estimate the uncertainty surrounding these parameter estimates to guide sensitivity analysis, where models are reanalyzed as certain key parameters are varied throughout a reasonable range of values. The particular parameters selected and the range through which they are varied is guided by knowledge regarding the range of reasonable estimates for these parameters, suspicion that their variation may substantially affect results, or a desire to explore the effect that potential changes in diagnostic accuracy, treatment effectiveness, or cost might have on predicted results.

In the evaluation of medical interventions, CEA use is based on the premise that decision makers would like to maximize health outcomes at any given level of spending (59,69). CEA helps to define the opportunity cost of selecting one intervention rather than another. Different options are compared by using comparable measures of cost and outcome, and the resulting incremental cost-effectiveness ratios can be used to determine the cost of each additional unit of health outcome. By funding programs in order of increasing incremental cost-effectiveness ratio (59), it is possible to obtain the maximum health benefits given a fixed budget. CEA, as seen in this manner, is a tool with which to optimize resource allocation to programs that compete for funds from the same limited budget.

CEA has been used to examine a variety of medical technologies, procedures, and programs. A review of these analyses, or the scope of the possible applications of CEA in medical technology assessment, is beyond the scope of this article. We direct interested readers to an online registry of CEAIs maintained by the Harvard Center for Risk Analysis (www.hsph.harvard.edu/cearegistry/).

The information provided by CEs can be used in many different settings. It can be used to guide policy making at the societal level, program implementation at the local or state level, purchasing decisions at the hospital or health system level, coverage decisions by third-party payers, or pricing and marketing decisions by pharmaceutical and device manufacturers.

Why, then, is CEA not universally used to determine how we spend all of our health care dollars? To begin with, cost-effectiveness is not the only important criterion in making decisions concerning resource allocation, nor should it be. Many additional factors are important when setting funding priorities. Simple mechanical optimization of limited resources by using cost-effectiveness ratios ignores important issues such as distributive justice, equity, and benefits and costs outside of the health care system (53). It also may not permit appropriate comparison of interventions that affect different subgroups of the entire population. For example, questions have been raised about comparing interventions that affect young people with those that affect old people, those that affect healthy people with those that affect poor people, or those that affect healthy people with those that affect unhealthy people. In each case, cost-effectiveness ratios may not be the most appropriate metric with which to compare interventions.

THE ROLE OF DECISION ANALYSIS

Decision analysis is an analytic approach that can be used to guide decision making when conditions of uncertainty exist (70–75). It allows the analyst to integrate all available information and take into account any underlying uncertainty or complex sequencing of events and decisions. Decision analysis has been used in a wide variety of settings, from stock market investment to cancer chemotherapy. It is particularly well suited for medical interventions, however, because of the complexity of clinical decisions and because virtually all clinical decisions are made in the face of some uncertainty. Moreover, decision analysis has been performed in different areas of medicine, such as primary prevention (risk factor reduction), secondary prevention, screening, diagnostic procedures, therapy, and rehabilitation (76). Several situations in which decision-analytic modeling is helpful or even required have been described elsewhere (77). In most of these situations, decision models are used for combining or linking data from different research areas and sources or transferring or extrapolating results from one time, place, population, or setting to another (77).

Decision analysis proceeds in a step-wise fashion. First, the problem to be addressed is identified and bounded. Next, the problem is structured, and the information that will be needed to analyze it is identified. Finally, the optimal course of action, based on the available data and analysis performed, is determined (72). Often there is no single best decision, or the answer is highly dependent on the particular assumptions of the decision model. In either case, the optimal decision must be qualified; however, if no single best choice can be determined, decision analysis can be a valuable tool in the identification of critical factors and their influence on the optimal choice. It can also be used to guide prospective data collection and efficiently narrow the bounds of uncertainty surrounding those variables or assumptions that have the greatest effect on the ultimate decision (78,79).

The use of decision analysis to guide complex resource allocation decisions has one important advantage: It forces individuals to be explicit about the factors that influence their judgments and decisions. The transparency provides a tool with which to assess the soundness of the judgments made based on the available data and to better understand the effect that changing assumptions might have on optimal choices. When disagreements exist, it is often possible to move toward consensus by being explicit about each component of the decision (eg, data, beliefs, event probabilities) and resolving disagreements about specific components one at a time rather than focusing on the decision at a macro level.
UNIQUE FEATURES OF DIAGNOSTIC TECHNOLOGY ASSESSMENT

Assessment of diagnostic technologies differ from the evaluation of medical therapeutics in many respects. One of the most important and challenging differences is the indirect relationship between the results of a diagnostic test and the actual health outcome in patients. Diagnostic test results are intermediate outcomes; they influence, but do not directly determine, health outcomes in patients. A number of factors can alter the usual chain of events and prevent the patient from receiving the intended treatment. For example, the physician may misinterpret the patient’s symptoms at presentation and order the wrong diagnostic test. The diagnostic test may be technically inadequate, or it may be technically adequate, but the specialist may err when interpreting the study. The correct interpretation may not be accurately transmitted to the referring physician, or the referring physician may select the wrong treatment. The patient may elect not to comply with the recommendations of the physician. Primarily depending on the severity of the disease and the effectiveness of the treatment used, the patient’s health may improve, stay the same, or deteriorate.

Although imaging procedures share the generic features of all diagnostic tests, there are several issues that are peculiar to imaging tests. First, the test results are often multidimensional (eg, presence or absence, location, size, form, and constitution of a tumor) rather than one-dimensional, as is the level of a tumor marker in a blood test. Second, clear cut points are rarely established; thus, test results must often be summarized in terms of likelihoods, such as “very unlikely,” “unlikely,” “likely,” and “very likely,” instead of in terms of categories, such as “test positive” and “test negative.” Third, images can reveal signs for different diseases, further adding to the complexity of the decision making process. Fourth, imaging techniques may be associated with the risk of radiation-induced side effects, leading to a clinical trade off between benefit and harm. Fifth, image quality increases with improved resolution; thus, the results of diagnostic studies are often outdated when devices with better image quality emerge. Sixth, many emerging imaging tests are expensive. Finally, an important feature for the evaluation of a diagnostic procedure is that images can be assessed at different times and by different readers, which allows us to analyze intra- and interobserver agreement.

Although more attention has recently been focused on diagnostic technology assessment, assessing the efficacy and effect of diagnostic imaging has been a concern of radiologists and policy makers for many years. Pioneering work in this field was performed by Ledley and Lusted (80), Thornbury et al (81), Fineberg and Hlatl (19), Fineberg (34), Fineberg and colleagues (35), Abrams and McNeil (30,31,43), McNeil and colleagues (37), McNeil and Adelstein (82), and other researchers. During the past two decades, methodologic and applied research on diagnostic technology assessment has blossomed, and many important issues have been addressed (83–90).

Abrams and McNeil (43) suggested that the complexity of the relationship between diagnostic test results and actual health outcomes has led to the use of more accessible “process” measures. The literature is replete with studies that have been performed to assess tests that use parameters such as “lesion conspicuity,” number of correct diagnoses made, and number of unanticipated findings. Even more reports have simply described the frequency with which a particular finding is associated with a disease or group of diseases or reported the spectrum of findings associated with one disease or another. Relatively fewer reports have estimated the effect of imaging technologies on final health outcomes at either the patient or the societal level.

A HIERARCHICAL APPROACH TO ASSESSMENT OF DIAGNOSTIC EFFICACY

Against the backdrop of the introduction of CT, Fineberg et al (35) proposed a hierarchical conceptual framework for use in the evaluation and comparison of diagnostic imaging tests, elements of which had already been described in the literature (81,82). The approach of Fineberg et al (35), later adopted and modified by others (91–94), includes sequential evaluation of the following six levels of efficacy: (a) technology, (b) diagnostic accuracy, (c) diagnostic thinking; (d) therapeutic planning; (e) patient outcomes, and (f) society. Efficacy at higher levels is contingent on efficacy at all lower levels, but improvements at any given level do not necessarily translate into improvements at higher levels. Furthermore, a test with fixed efficacy at one level (N) may have different efficacies at higher levels (N + 1) in different clinical situations.

Level 1, or “technical efficacy,” generally falls within the domain of physicists and engineers who develop and refine an imaging technology before its clinical implementation and testing. Technical efficacy is usually judged by using parameters that can be precisely measured in a laboratory with optimal conditions, such as spatial resolution, quantum mottle, necessary exposure time, and radiation dose.

Level 2, or “diagnostic accuracy efficacy,” of a binary test is expressed by using sensitivity and specificity, positive and negative predictive values, or receiver operating characteristic (ROC) curves. The two most common measures of test performance are sensitivity, TP/(TP + FN), and specificity, TN/(TN + FP), where TP is the number of true-positive results, FN is the number of false-negative results, TN is the number of true-negative results, and FP is the number of false-positive results; these can be calculated when a reliable reference standard is available. Biases may result, however, when the reference standard is either imperfect or not uniformly applied (described in more detail later) or when test results are uninterpretable, distorted by measurement bias, or caused by differences in the “case mix” between study populations (95–97).

Often, the performance of new tests is evaluated by using existing diagnostic methods as a reference standard. If these existing diagnostic methods are imperfect (ie, “tarnished” reference standard), however, it may be impossible to determine whether what appears to be low specificity (ie, many false-positive results) represents truly poor specificity or improved sensitivity, relative to the reference standard. That is, the new test may enable the detection of abnormalities that were undetectable by using the standard of reference; thus, the results of the new test are incorrectly classified as false-positive interpretations. Attempts to improve the standard of reference by using subsequently obtained consensus readings (generally including all available diagnostic information) may yield biased estimates, particularly if the consensus panel treats a positive result with either test as indicative of true disease.

Additional problems may arise when the reference standard is not uniformly applied to all patients, and the probability of verification depends on the diagno-
The odds-likelihood ratio form of Bayes theorem provides a framework for evaluating the effect of diagnostic information on diagnostic thinking (72):

$$\frac{P(D^+|R)}{P(D^-|R)} = \frac{P(D^+)}{P(D^-)} \frac{P(R|D^+)}{P(R|D^-)}$$

or

posterior odds = prior odds \cdot likelihood ratio,

where $P(D^+)$ represents the probability disease ($D$) is present (prevalence), $P(D^-)$ represents the probability that disease is not present, $P(D^+|R)$ represents the probability of disease given the test result ($R$), $P(D^-|R)$ represents the probability of no disease given the test result, $P(R|D^+)$ represents the probability of result given the disease, and $P(R|D^-)$ represents the probability of result given no disease.

The following equation was used to calculate the prior odds favoring ($O_p$) disease being present:

$$O_p = \frac{P(D^+)}{P(D^-)}$$

The following equation was used to calculate the prior odds against ($O_a$) disease being present:

$$O_a = \frac{P(D^-)}{P(D^+)}$$

Similarly, the following equation was used to calculate the posterior odds favoring disease being present:

$$O_p' = \frac{P(D^+|R)}{P(D^-|R)}$$

Finally, the following equation was used to calculate the posterior odds against disease being present:

$$O_a' = \frac{P(D^-|R)}{P(D^+|R)}$$

The likelihood ratio represents the ratio of the frequency of a certain test result in patients with disease to its frequency in patients without disease. The likelihood ratio can be used to judge the usefulness of a particular test in a given clinical situation (81,106–108). Advocates of evidence-based medicine have also recommended the use of likelihood ratios in the evaluation of diagnostic technologies (109,110).

Level 4 is known as “therapeutic planning efficacy.” The greatest efficacy at this level results from a test that might (correctly) lead to the initiation of a new therapy or the determination that therapy is not required. Studies concerned with level 4 efficacy are generally performed to compare intended patient care strategies prior to the test with intended patient care strategies after the test. Studies of the efficacy of body CT by Wittenberg and colleagues (32,33) illustrate this type of assessment. Level 4 efficacy questions are extremely challenging, however, since it is often difficult to determine what would have been done without the results of a diagnostic test once those results are available. Even if one were to query physicians prior to providing them with the test results, the situation is artificial; they know that the test has been performed, and neither they nor their patients will face any real risks on the basis of answers to these hypothetical questions.

Level 5, or “patient outcome efficacy,” can really only be assessed in a prospective RCT, in which some patients undergo the test but others do not, and patient outcomes in the two groups (test vs no test) are compared. Unfortunately, RCTs of this sort are difficult to perform and are associated with challenging practical, analytic, and ethical issues. Imaging tests may have high sensitivity and specificity, cause important changes in clinicians’ diagnostic thinking, and even cause different therapies to be instituted. If these changes cannot be translated into improved patient outcomes, however, the value of the test—at least from the patients’ perspective—is questionable.

An additional challenge is to combine patient outcomes in different dimensions (eg, years of life gained, quality of life, preference for one test over another) into a single meaningful outcome measure (46–48,111,112). For example, in some cases, the psychologic effects of diagnostic information, whether positive or negative, may be more important to patients than other effects that the test might have on outcomes. Examples are the re-assurance value of a negative test (113) or the anxiety introduced by an imperfect or ambiguous test result. Determining how to relate these effects to overall quality of life remains to be investigated.

Level 6, or “societal efficacy,” asks whether the benefit to society associated with the use of a test is acceptable in relation to its cost. In other words, is the test an efficient use of societal resources? In the current era of cost-conscious medical care, no major technology assessment effort should be considered complete without addressing cost-effectiveness considerations; however, these assessments are often the most challenging to perform. Level 6 assessments share all of the difficulties related to the assessment of efficacy of level 1–5 assessments, and they require a substantial amount of time and analytic resources to complete. A further complication is estimation of the cost of
newly developed tests. Prices of new or experimental imaging technologies may not adequately represent their true cost in a competitive marketplace, and cost-based reimbursement rates often do not yet exist or are woefully inaccurate. Conducting a detailed analysis of each component of a particular test or procedure can begin to address these issues, but this analysis cannot possibly be performed for each new test as it becomes available. Finally, most new imaging technologies continue to undergo rapid technologic evolution in the years immediately after clinical implementation. Even if it were possible to estimate the costs and effects of each new test, it would be unreasonable to expect the estimates to be accurate for long. All too often, level 6 analyses cannot be accomplished before decisions regarding resource allocation must be made.

When deciding whether or not to commit resources to new imaging technologies, physicians have rarely considered costs and effects at a societal level. There are a number of reasons for this. To begin with, neither physicians nor patients directly bear the costs of diagnostic testing or therapeutic medical interventions. In addition, decisions are generally made in the context of individual patient encounters. In this context, it is difficult for a physician to withhold a test or treatment that may provide some benefit. It is even more difficult for the patient—who may be facing a life-threatening illness—to accept that a test or therapy is withheld because of concerns about optimizing resource allocation at a societal level.

**DECISION ANALYSIS AND DIAGNOSTIC TECHNOLOGY ASSESSMENT**

A decision analytic model (72,114–118) can be used to integrate the data from all of these hierarchical levels of efficacy and evaluate final patient outcomes in relation to diagnostic test results, across a range of varying assumptions concerning disease prevalence, test performance, therapeutic efficacy, and costs. It is possible to model patient outcomes beyond the time horizons of short-term treatment studies. Sensitivity analysis can then be used to analyze the effect of varying assumptions on the conclusions reached and determine the conditions in which a particular test or combination of tests would be optimal.

Phelps and Mushlin (117) developed a strategy for evaluation of diagnostic technologies by using medical decision theory. The Phelps and Mushlin model presumes that there is some societal threshold, in dollars per quality-adjusted life-year, above which the investment of societal resources cannot be justified, and below which it is generally not questioned. Given ROC parameters for an existing procedure, a “challenge region” is computed and projected into ROC space. The ROC curve for a new imaging test must fall into the challenge region for its incremental cost-effectiveness to be less than the societal threshold.

The analysis proceeds in two steps. First, expected costs and patient outcomes with the new test (assuming perfect diagnostic accuracy) are compared with expected costs and outcomes in the absence of the additional diagnostic information that the new test might provide. If, by using these most optimistic assumptions, the ROC curve of the new test fails to enter the challenge region, it can be eliminated from further consideration (hurdle 1). If this first hurdle is met, clinical studies are undertaken to more precisely define the performance characteristics of the new test; the decision model can help identify critical information to be assessed. Next, by using the results of the clinical trials in combination with the decision models already developed, the incremental cost-effectiveness of the new test (relative to the existing alternative) is recalculated and compared to the (hypothetical) societal acceptability threshold (hurdle 2).

The principal advantage of the Phelps and Mushlin (117) analytic approach is that societal cost-effectiveness can be addressed by using relatively straightforward level 2, or diagnostic accuracy, efficacy data. By performing the assessments in two steps, tests that are “noncontenders” may be eliminated prior to determining sensitivity and specificity and constructing ROC curves. For those tests that pass the first hurdle (i.e., their ROC curves fall into the challenge region under the assumption of perfect diagnostic accuracy), the decision models can help focus prospective data collection to ensure that necessary data are collected efficiently. The use of medical decision theory may permit more costly and time-consuming clinical trials to be either avoided altogether or streamlined substantially. If the decision to implement a new test would not vary over a range of possible values of some parameters, then expensive and time-consuming data collection may be avoided (78,79).

**RANDOMIZED CONTROLLED TRIALS IN DIAGNOSTIC TECHNOLOGY ASSESSMENT**

Decision analytic methods necessarily involve the use of simplifying assumptions and may rely on inaccurate parameter estimates. Nonrandomized trials or retrospective data collection may produce overly optimistic estimates of benefit. RCTs are often required to confirm the results of a decision analysis and to be certain that confounding has not biased the results in favor or against the test being considered.

Designing and conducting an RCT to assess diagnostic imaging tests is less straightforward than studying therapies. In an RCT of a therapy, the treatment is generally withheld from some patients and provided to others. However, because most diagnostic tests have minimal or no adverse effects, and because many patients and physicians may believe that more information results in better care, it is often difficult to withhold a diagnostic test from patients enrolled in a trial. It might be possible to perform both tests in all patients and only randomize those patients with discordant test results; however, even that creates ethical dilemmas.

In an alternative approach (94), both tests are performed in all patients, but one of the tests is chosen at random as the test that will be used to determine therapy. Only the results of the randomly chosen test are provided to the physician(s) caring for the patient. By using this study design, it may also be possible to determine—in a separate setting—what care decisions would have been made if the results of a test that was not selected had been used. The advantage of this approach is that long-term follow-up may be omitted for patients with concordant (and in some cases, discordant) test results because the same care decisions would nevertheless have been made. This approach can only be used if the tests do not have complications that result in long-term consequences and if all of the possible effects of each test can be distinguished from one another and from the effects of therapy.

With the ever-increasing number of diagnostic tests available, it is often the case that several tests or combinations of tests are potentially useful. Comparing all of these tests in a clinical trial with sufficient statistical power to detect meaningful differences in outcome may require an extremely large trial and could
Radiology

May 2005

Gazelle et al

368

 Radiology facing modern medicine. we continue to address the challenges and interventions will be required as the costs and benefits of medical technologies, makes their use in as well as the continuous evolution of medical expertise to actually weigh benefits and harms their patients); however, the re- cipient of technology assessment and re- strict the development and clinical im- plementation of expensive new thera-pies. A truly cost-sensitive financing system, he argues, will encourage providers to avoid overly expensive technolo- gies and discourage researchers and manufacturers from developing them. Medical care in the United States is pro- vided in an absolutist environment, however, and both professional and legal standards rarely permit an individual physician or payer to withhold poten- tially beneficial treatment from patients. Providers certainly do have appropriate incentives to avoid treatments with no benefit (and especially those that might harm their patients); however, the re- quirements for technology assessment in this setting are more modest (i.e, to provide information regarding efficacy alone). The use of technology assessment in a broader role is unlikely because regulators lack the expertise to actually weigh benefits and costs and because the technical complexity of medical technology assessments, as well as the continuous evolution of med- ical technologies, makes their use in ac- tual practice difficult.

Despite the valid concerns, it seems clear that more careful consideration of the costs and benefits of medical technol- ogy and interventions will be required as we continue to address the challenges facing modern medicine.

LIMITATIONS OF TECHNOLOGY ASSESSMENT

There is little question that technology assessment of medical technologies and practices is here to stay. The more critical questions are as follows: What can we hope to gain from increased efforts in this area? What will be the challenges that lie ahead?

Elhauge (119) has argued that technol- ogy assessment is important in our quest to better understand and improve the quality of medical care, but technology assessment is relatively limited in its ability to address the specific problem of cost escalation. He argues that if we are to constrain the rapid increase in health care spending, proper incentives to trade off costs and benefits are needed, rather than more technology assessments. He further suggests that, if anything, the move to a more cost-sensitive means of health care financing is likely to decrease the use of technology assessment and re- strict the development and clinical im- plementation of expensive new thera- pies. A truly cost-sensitive financing system, he argues, will encourage provid- ers to avoid overly expensive technolo- gies and discourage researchers and manufacturers from developing them. Medical care in the United States is pro- vided in an absolutist environment, however, and both professional and legal standards rarely permit an individual physician or payer to withhold poten- tially beneficial treatment from patients. Providers certainly do have appropriate incentives to avoid treatments with no benefit (and especially those that might harm their patients); however, the re- quirements for technology assessment in this setting are more modest (i.e, to provide information regarding efficacy alone). The use of technology assessment in a broader role is unlikely because regulators lack the expertise to actually weigh benefits and costs and because the technical complexity of medical technology assessments, as well as the continuous evolution of med- ical technologies, makes their use in ac- tual practice difficult.

Despite the valid concerns, it seems clear that more careful consideration of the costs and benefits of medical technol- ogy and interventions will be required as we continue to address the challenges facing modern medicine.

References


