The Efficacy of Diagnostic Imaging

DENNIS G. FRYBACK, PhD, JOHN R. THORNBURY, MD

The authors discuss the assessment of the contribution of diagnostic imaging to the patient management process. A hierarchical model of efficacy is presented as an organizing structure for appraisal of the literature on efficacy of imaging. Demonstration of efficacy at each lower level in this hierarchy is logically necessary, but not sufficient, to assure efficacy at higher levels. Level 1 concerns technical quality of the images; Level 2 addresses diagnostic accuracy, sensitivity, and specificity associated with interpretation of the images. Next, Level 3 focuses on whether the information produces change in the referring physician's diagnostic thinking. Such a change is a logical prerequisite for Level 4 efficacy, which concerns effect on the patient management plan. Level 5 efficacy studies measure (or compute) effect of the information on patient outcomes. Finally, at Level 6, analyses examine societal costs and benefits of a diagnostic imaging technology. The pioneering contributions of Dr. Lee B. Lusted in the study of diagnostic imaging efficacy are highlighted. *Key words:* diagnostic imaging; efficacy studies; cost–effectiveness; ROC analysis. **(Med Decis Making 1991;11:88–94)**

It is clear that the appraisal of outcomes of care will have a major impact on medicine in the 1990s. Relman[1] has called this the "third revolution in health care." Health care purchasers are demanding an accounting of value received for their dollars spent. The technology assessment community is attempting to provide the analytic infrastructure to make this accounting. In this context, we discuss in this paper a conceptual model to organize the assessment of efficacy of diagnostic imaging.

Medical imaging is a major activity in modern medicine. Projections for the 1990s are difficult to find, but past utilization figures show diagnostic imaging to consume a significant portion of the medical budget. Diagnostic imaging is a major medical activity in the United States. It is estimated that the radiation used in medical and dental diagnoses contributes approximately 15% of the average annual effective dose equivalent to individuals in the U.S. population from all sources. Diagnostic imaging is clearly the largest man-made component of population radiation exposure, second only to natural background exposure.[2,3] In 1980 the annual costs of medical diagnostic imaging were estimated at $5–7.5 billion.[4] Moreover, the number of hospital-based diagnostic imaging examinations increased by about 60% between 1970 and 1980.[5] With the increasing use of diagnostic imaging methods, there is wide variation from one section of the United States to another in the frequencies with which imaging examinations are requested by physicians.[6] These data as well as issues such as unnecessary radiation exposure, disagreement over the clinical value of specific examinations, and concern about the expenditure of scarce medical resources, have stimulated interest in studies aimed at assessing the usefulness of various imaging procedures and examinations.

Assessing the efficacy of diagnostic imaging has long been a concern of the radiologic community. In the preface to his 1968 book on medical decision making, Lusted[7] notes his participation as a subject in a 1954 study by Yerushalmy about observer error in radiologic diagnosis. That study[8] could well be considered an efficacy study in diagnostic radiology, as it appraised one component of accuracy in diagnostic imaging. That experience, and further thought and collaboration about the probabilistic nature of radiologic diagnosis, led to the paper by Ledley and Lusted[9] that is cited as the root of the "evolutionary tree for medical decision making."[10] From its beginnings in Bayesian approaches to the problem of medical diagnosis, medical decision making as a field has changed considerably. Diagnosis is now seen as only one part of the decision process, and we have sophisticated models for therapeutic decision making. Similarly, the intellectual tools of medical decision making, grown larger yet, can contemplate the cost–effectiveness ratios of alternative health policies.

Conceptual thinking about, and assessment of, efficacy of diagnostic imaging have evolved on a path similar to that of medical decision making. The various approaches to assessing efficacy along this path can be placed in a framework of medical decision making as a conceptual model for efficacy assessment. Our purpose in this paper is to elaborate this model.

"Efficacy" is defined in dictionaries as "effectiveness and efficiency." In the 1970s the concept of clinical efficacy began to crystallize. Work by Cochrane in 1972,[11] Thornbury et al. in 1975,[12] Fineberg in 1978,[13] and Lusted

Received June 11, 1990, from the Department of Preventive Medicine (DGF) and the Department of Radiology, Clinical Sciences Center (JRT), University of Wisconsin–Madison, Madison, Wisconsin. Accepted for publication June 11, 1990.

Address correspondence and reprint requests to Dr. Fryback: University of Wisconsin–Madison, 1513 University Avenue, Madison, WI 53706

Downloaded from http://mdm.sagepub.com at ALBERTA UNIVERSITY on March 14, 2008
© 1991 Society for Medical Decision Making. All rights reserved. Not for commercial use or unauthorized distribution.

et al. in 1980[14] did much to provide a conceptual framework for evaluating and comparing the efficacies of diagnostic tests. More broadly, the American College of Physicians established an initiative to study and disseminate information about efficacy of tests and procedures in the early 1980s.[15] More recently, Brook and Lohr[16] at the Rand Corporation in 1985, and Guyatt et al.[17] in 1986 at McMaster University have provided further conceptual and clinical underpinnings for efficacy studies. Since 1980, efficacy studies have been undertaken with increasing frequency. This has been spurred by the increasing rate of development of new and competing imaging technologies, the latest of which is magnetic resonance imaging.

Just as the concepts of efficacy have evolved over time, so have definitions and terminology. We take efficacy to be defined generically as "the probability of benefit to individuals in a defined population from a medical technology applied for a given medical problem under ideal conditions of use."[16] In the case of diagnostic technologies, the term "efficacy" overlaps a good deal with the idea of "effectiveness." By contrast, the term "effectiveness" reflects performance of a medical technology under ordinary, rather than ideal, conditions.[16] Other related terms that have been used at one time or another include "usefulness," "cost–benefit," and "cost–effectiveness."

We describe here a general hierarchical model of diagnostic imaging efficacy that can be used for classification of assessment studies. This hierarchy extends from basic laws of physics, through practical clinical use, to more general patient outcome and societal issues. Radiology, which has a long history of self-evaluation, represents one of the most thoroughly studied areas vis-à-vis clinical efficacy. Even so, it will become apparent that much remains to be done—particularly with regard to the higher efficacy levels.

A Model of Efficacy

The efficacy model presented here departs from a more traditional view. The differentiation is characterized by a change in perceived goals. A localized view of the goal of diagnostic radiology would be that it should provide the best images and the most accurate diagnoses possible. Certainly this should be a prime concern of radiologists. But a more global analysis reveals diagnostic radiology to be part of a larger system whose goal is to treat patients effectively and efficiently. In this larger context, even high-quality diagnostic imaging may be noncontributory in certain instances, and radiology of lesser quality may be of great value in others. The key thing to note is that the more global, "systems" view of efficacy forces one to consider standards that go beyond quality or accuracy of radiologic examinations to examine the ultimate value or benefit that is derived from those examinations.

This model is not new. Elements of it have been described in numerous places—Thornbury et al.,[12] first explicated Levels 3, 4, and 5. McNeil and Adelstein[18] also addressed elements of the model. Fryback[19] initially described this as a six-tiered model. Fineberg introduced an entire issue of *The American Journal of Roentgenology* devoted to efficacy of imaging.[13] He described efficacy research on computed tomography through all six levels. Fineberg's editorial marks the first discussion in the literature of the complete continuum of efficacy. The model concept as put forth in the present paper was developed in 1982 by a working committee for the National Council on Radiation Protection and Measurements (see acknowledgement).

THE SIX-TIERED MODEL—A CONCEPTUAL CONTINUUM FOR EFFICACY

The production and use of information from diagnostic imaging is a process with a number of parts controlled by several actors. The imaging process depends on an imaging device that records images in some medium. The image is, in turn, commonly interpreted by a radiologist, who produces a written (and possibly verbal) report to the clinician. The imaging process is embedded in a clinical process, whereby a clinician selects the patient for the procedure and uses its result as information for clinical decision making. Further, this whole system is embedded in a larger health care system in which some accounting of aggregate benefit to society as a whole may influence what diagnostic imaging resources are available for the clinician to select.

By the nature of the imaging process and the context in which it takes place, diagnostic imaging efficacy measures may be thought of as being arranged hierarchically (table 1). At the most micro, or local, level, we are concerned with information transmission measures describing the physical imaging process itself. We call this Level 1, and denote it "technical efficacy" of diagnostic imaging. Level 2, "diagnostic accuracy efficacy," includes the person(s) interpreting the image as well as the images per se. Level 3, "diagnostic thinking efficacy," addresses the proximal effect on the clinician who ordered the test, and Level 4, "therapeutic efficacy," a slightly more distal point of effect on the clinician's choice of therapy. Level 5, "patient outcome efficacy," is concerned with whether there is measurable effect of the image on the outcome experienced by the patient. Finally, a sixth level, "societal efficacy," defines the most global level of efficacy.

This hierarchical model can be applied with equal ease to nearly all diagnostic technologies—not just diagnostic imaging technology. The history, physical examination, biochemical tests, even the "test of time," all qualify as diagnostic technologies. Use of this model may, in fact, facilitate comparison of the various "technologies" (e.g., imaging versus physical examination).

Downloaded from http://mdm.sagepub.com at ALBERTA UNIVERSITY on March 14, 2008
© 1991 Society for Medical Decision Making. All rights reserved. Not for commercial use or unauthorized distribution.

Table 1 • A Heirarchical Model of Efficacy: Typical Measures of Analyses

Level 1. Technical efficacy
Resolution of line pairs
Modulation transfer function change
Gray-scale range
Amount of mottle
Sharpness
Level 2. Diagnostic accuracy efficacy
Yield of abnormal or normal diagnoses in a case series
Diagnostic accuracy (percentage correct diagnoses in case series)
Predictive value of positive or negative examination (in a case series)
Sensitivity and specificity in a defined clinical problem setting
Measures of ROC curve height (d') or area under the curve A_z
Level 3. Diagnostic thinking efficacy
Number (percentage) of cases in a series in which image judged "helpful" to making the diagnosis
Entropy change in differential diagnosis probability distribution
Difference in clinicians' subjectively estimated diagnosis probabilities pre- to posttest information
Empirical subjective log-likelihood ratio for test positive and negative in a case series
Level 4. Therapeutic efficacy
Number (percentage) of times image judged helpful in planning management of the patient in a case series
Percentage of times medical procedure avoided due to image information
Number or percentage of times therapy planned pretest changed after the image information was obtained (retrospectively inferred from clinical records)
Number of percentage of times clinicians' prospectively stated therapeutic choices changed after test information
Level 5. Patient outcome efficacy
Percentage of patients improved with test compared with without test
Morbidity (or procedures) avoided after having image information
Change in quality-adjusted life expectancy
Expected value of test information in quality-adjusted life years (QALYs)
Cost per QALY saved with image information
Level 6. Societal efficacy
Benefit–cost analysis from societal viewpoint
Cost–effectiveness analysis from societal viewpoint

A key feature of this model is the understanding that for an imaging procedure to be efficacious at a higher level in this hierarchy it must be efficacious at lower levels, but the reverse is not true. Increases in the efficacy at a lower level (e.g., technical image quality) will not guarantee commensurate improvement at higher levels (e.g., patient outcome). This asymmetry is often lost to view in research reports at Levels 1 and 2.

We discuss briefly each of the six levels of efficacy.

Level 1, Technical Efficacy. Technical efficacy of diagnostic imaging is generally the purview of physicists who are concerned with the physical parameters describing technical image quality in an imaging system. These include the modulation transfer function (MTF), sharpness, brightness, contrast, mottle, and time required for an optimal exposure. Such parameters are usually derived under optimal laboratory conditions and are prerequisites for consideration of efficacy at subsequent levels.

Additional important variables in obtaining high-quality images include the presence of artifacts related to the patient or from the imaging itself, as well as the ability of the radiologic technologist to produce high-quality images. Because the volume of literature in this area is so large and because technical efficacy is the most conceptually well developed and well understood level of imaging efficacy, this level is not reviewed in this overview.

Level 2, Diagnostic Accuracy Efficacy. Diagnostic efficacy has been characterized by such measures as number of abnormalities found in a case series (yield), accuracy of diagnosis (often measured as percentage correct in a case series), predictive values for positive and negative examinations, sensitivity, specificity, and measures relating to receiver operating characteristic (ROC) curves such as d' (a measure of discrimination between normals and abnormals afforded by the test) and A_z (the area under the ROC curve). Important to all these measures are that they attempt to measure performance of the imaging for the purpose of making diagnoses and that they all require interpretation of the image by an observer.

While it is still possible to find in the literature simple studies counting the number of abnormal patients found in a case series undergoing a particular diagnostic imaging test (e.g., many of the studies cited in a review of magnetic resonance imaging[20]), more sophisticated concepts of test performance are becoming more prevalent in the literature evaluating imaging tests.[21] The signal-detection paradigm, which utilizes the receiver operating characteristic curve, is a particularly useful means of characterizing diagnostic imaging examination performance because it explicitly acknowledges the role of the observer.[22,23] Among the first to recognize the usefulness of ROC curves in characterizing performance in diagnostic imaging was Lusted.[7,24,25] A number of advances in thinking about derivation and analysis of ROC data have appeared in *Medical Decision Making*,[26–29] so we do not discuss this topic more deeply here.

All of these measures illustrate an important concept: diagnostic accuracy efficacy is not simply a function of the image. It is a *joint* function of the images and of an observer, such as a radiologist. Only very recently has it been discussed in the literature that Level 2 diagnostic imaging efficacy is also a function of the clinician who requests diagnostic imaging, as it is this selection that controls both the specificity of a test in the clinical practice environment[30] and the sensitivity of the test to the extent that it varies with spec-

Downloaded from http://mdm.sagepub.com at ALBERTA UNIVERSITY on March 14, 2008
© 1991 Society for Medical Decision Making. All rights reserved. Not for commercial use or unauthorized distribution.

trum of disease. And, a most vexing problem is how to establish the "true" diagnosis in an ROC experiment, the "gold standard" problem.[21,31]

Clearly, image quality, i.e., technical efficacy, contributes to diagnostic accuracy efficacy. But it also becomes apparent that there may be a point beyond which improvement in technical efficacy no longer improves diagnostic accuracy efficacy. This illustrates the essential asymmetry in relationships between adjacent levels in the continuum of efficacy.

In spite of the problems of spectrum in control patients[30] and of the problems with "gold standard" diagnosis,[31] ROC analysis of diagnostic imaging tests is relatively feasible empirically.[23] Taking the next step, and showing an imaging test to contribute to an improved patient outcome, is much more difficult. Assurance of the diagnostic accuracy of an imaging test does not guarantee it will in fact contribute to improved management of patients.

Level 3, Diagnostic Thinking Efficacy. The impact of the diagnostic imaging information on the diagnostic thinking of the clinician who ordered the test was proposed as an intermediate step linking the information content of the image to changes in the treatment of the patient.[12] Measuring change in diagnostic thinking was used as an empirically feasible proxy for measuring ultimate impact on the patient. It is exceedingly difficult, if not impossible, outside of a prospective randomized trial to attribute some portion of improved patient condition to the use of an imaging test. Rarely can such a trial be arranged—especially for tests already in common practice. But by reasoning backwards we can see that inducing change in the clinician's diagnostic thinking is a necessary prerequisite to having impact on patients. In a particular clinical problem, patients' outcomes cannot be affected by the image information unless the attending clinician is led to do something different than would have been done without the information. Similarly, the physician's choice of management should not change unless something has changed in the diagnostic thinking, all other things equal. The imaging information may change the differential diagnosis, strengthen an existing hypothesis, or simply reassure the physician. Reassurance here usually means reassurance that the patient does not have an occult, unexpected serious disease present.

Although we can *compute* the normative pretest to posttest change in the probability of a disease if we know the imaging examination's sensitivity and specificity,[32] whether in the clinical environment the examination changes the referring clinician's subjective probability of the disease is an *empirical* question. It is possible in principle to observe impact of diagnostic information on clinicians' diagnostic thinking using subjectively estimated diagnostic probabilities. Clinicians are asked to identify one or more of the diagnoses in the differential (e.g., the most likely one, or the most important or threatening one) before imaging information is given to them. For each diagnosis they estimate, based on all information to that point, the probability that the diagnosis will turn out to be the true cause of the patient's problem. Then the image information (the radiologist's report or the report and the images) is presented and the question is repeated. Although there are problems with interpreting exact magnitudes of subjectively estimated probabilities,[33] this method can identify when no change in probabilities results from the information, and situations in which the diagnostic information is unlikely to affect decisions can be identified.[12]

The image may have non-decisional impact as well. It is not uncommon to see clinicians place great value on results that do nothing more than reassure them. As such, measuring the impact of the imaging result on clinician thinking can be a complex and value-laden exercise. Because questions of efficacy are often raised in the context of asking what impact the dollars spent on imaging have on the patient (as opposed to the clinician), the most useful studies are those that directly measure the impact of new information (i.e., the imaging result) on the physician's subjective diagnostic probabilities.[12,14,34] At this level, efficacious imaging examinations are those that significantly raise or lower pretest diagnostic probabilities assumed by the physician. Because of the subjective nature of the estimates and multiple sources of bias, the empirical methods are probably best for determining the *absence* of diagnostic thinking efficacy rather than estimating the magnitude of change in diagnostic thinking due to imaging information.

Level 4, Therapeutic Efficacy. An imaging examination result may influence the physician's diagnostic thinking and yet may have no impact on patient treatment. The most efficacious studies at level 4, of course, are those that lead to the institution of new therapy or else avert the need for therapy. Conversely, imaging examinations that have no impact on therapy cannot be expected to benefit the patient, except by means of reassurance of the physician. Some would argue (as in Level 3) that the reassurance of the physician's pretest treatment plan has an efficacious effect in itself through maintenance of the optimum therapy approach.

The empirical questions asked by Level 4 studies concern comparison of intended management (before the diagnostic examination is obtained) with the actual treatment pursued (after the results are known). Did the imaging information change the a priori course of management? Did the procedure affect management favorably?

The empirical questions are difficult to examine as they require either prospective assessment of therapeutic intent before and after the examination or ask-

Downloaded from http://mdm.sagepub.com at ALBERTA UNIVERSITY on March 14, 2008
© 1991 Society for Medical Decision Making. All rights reserved. Not for commercial use or unauthorized distribution.

ing the hypothetical question, "What would you have done for the patient if the diagnostic examination were not available?" In certain special situations, an effect might be inferred retrospectively from the clinical records—e.g., in patients already scheduled for surgery, it is possible to examine whether the surgery is canceled because of a preoperative x-ray result.[35,36] The validity of answers to these questions is controversial. In addition, the logistics of performing pre- and posttest assessments are complicated and difficult to carry out in the study of daily clinical practice. In situations where prospective randomized trials of decision making with and without the imaging information cannot be performed ethically or because of the momentum built for using a particular procedure, asking the questions at Level 4 may be the only efficacy study possible. This type of efficacy study is particularly well illustrated by Level 4 studies of computed tomography.[37,38] Integrating negative information about a test from Level 3 and 4 studies can help to direct clinical use away from imaging examinations that are not useful, or have been supplanted by other examinations.[39]

Level 5, Patient Outcome Efficacy. Of course, the important ultimate goal of medical care has been to improve, or return to normal, the health of the patient. It is possible for a diagnostic examination to appear efficacious for use in a particular patient population in every other way but have no appreciable effect on patient outcomes. For imaging examinations that are expensive, dangerous, or widely used, knowledge about patient outcome efficacy seems particularly important.

Level 5 is the sine qua non of efficacy from the individual patient's viewpoint. The patient outcome study is the first point at which the expected costs (radiation exposure, monetary, pain, risk to life, etc.) of an examination may be directly weighed against its expected benefits (improving life expectancy, improving quality of life, avoiding other tests and procedures, etc.) as a rational guide for the clinician's decision about whether or not to obtain that examination.

A definitive answer concerning whether a radiologic examination is efficacious with respect to patient outcome requires a prospective, randomized controlled trial (RCT). Such a trial requires withholding the test from some of the patients. The statistical, empirical, and ethical problems associated with such a trial can be formidable and may be justified only in carefully selected circumstances. In fact, there have been several such studies conducted, for example, that of Marton et al.[40] In many circumstances the question of patient outcome efficacy is approached by case series collected before an imaging technique was available and after it was available, or case–control studies. The limitations of such studies for inference of efficacy are well known, and often weaken considerably the conclusions concerning patient outcome efficacy. Further limiting the ability of empirical study of patient outcome efficacy is the fact that imaging information is not the only basis for decision making about management of patients. The independent contribution of imaging to patient outcomes may be small in the context of all the other influences on patient outcomes. To determine empirically the unique contribution of imaging can thus require very large sample sizes.

In the absence of a direct empirical approach, statistical methods using epidemiologic data may be used for attacking the problem of demonstrating patient outcome efficacy. Thus, the decision-analytic approach provides an investigative alternative to the traditional randomized controlled trial. One feature that makes decision analysis attractive for determining patient outcome efficacy for different imaging methods is that each patient can have both (or all) of the competing examinations (e.g., computed tomography and magnetic resonance imaging). Another advantage of the decision-analytic approach is that it is less expensive and requires less time than the randomized controlled trial. Two early examples of a decision-analytic approach to determining efficacy of an imaging procedure in a particular clinical problem are analyses by McNeil and colleagues,[41,42] whose conclusions appear to have withstood relatively well the effect of passage of ten to 15 years.[21] A more recent example in the same framework is an analysis of computerized tomography versus magnetic resonance imaging in dementia, appearing in *Medical Decision Making*.[43]

Although a decision analysis of whether to use a diagnostic imaging examination is in principle based on patient outcome efficacy, the actual analysis is pieced together from secondary data. Thus it may suffer from the same biases as before–after studies. While sensitivity analysis may compensate for some of these bias problems, it is still limited. Decision analysis can highlight which pieces of information are most critical and thus influence future study designs for determining patient outcome efficacy.

Level 6, Societal Efficacy. For the policy maker entrusted with the job of making resource allocations for large groups, the societal question of efficacy goes beyong the question of individual risks and benefits. Is the cost (borne by society as a whole) for use of a given examination acceptable, even though each individual patient application may be efficacious at some other level? At this level a diagnostic imaging examination is efficacious to the extent that it is an efficient use of societal resources to provide medical benefits to society.

The commitment of resources to diagnostic imaging appears to be rarely questioned at this level except for introduction of new technology. This happened when computerized tomography was introduced, promising imaging beyond the dreams of most at the time, but at a nearly unheard-of price. The introduction of mag-

Downloaded from http://mdm.sagepub.com at ALBERTA UNIVERSITY on March 14, 2008
© 1991 Society for Medical Decision Making. All rights reserved. Not for commercial use or unauthorized distribution.

netic resonance imaging was not quite the quantum change that CT represented in the early 1970s, but its very noticeably higher price put it in this category. Advocacy of widespread use of mammographic screening for breast cancer,[44–46] and the similarly large financial-impact potential of transrectal ultrasound screening for prostate cancer[47] have catapulted these procedures to national prominence for evaluation at Level 6. As an adjunct, the use of low-osmolality contrast media and the concomitant large increases in the costs of nonionic contrast-enhanced imaging studies have placed this innovation at the top of the Level 6 analysis list as well.

Many issues of analysis and policy are raised in facing Level 6 problems. These are well stated by Fineberg[13] with reference to CT:

> Evaluation of CT, or of any dynamic medical technology, will never provide final answers. Findings will be open to inperpretation. Individual values and judgments will always play a role. Decisions about the development, reimbursement, and use of new technologies will continue to be made, however imperfectly. The great challenge, embodied in CT but embracing all of medicine, is to bring policy and practice into line with knowledge. (p. 3)

The Relation of Efficacy to Cost–Effectiveness

Cost–effectiveness studies compute a cost (or price) per unit of output for a medical procedure or technology. Any of the measures in table 1 (with some work, perhaps) from Levels 1–4 could be used as an output measure for a cost–effectiveness analysis. For example, we could compute cost per surgery avoided, or per changed management plan, or per change in diagnostic thinking, or per correct diagnosis. Often the purpose of cost–effectiveness analysis is to compare resources invested in the particular treatment or technology with widely different investments in other modes of health care. For this purpose a more general measure of output is needed. Commonly this now is life years (LYs) saved, or quality-adjusted life years (QALYs) saved. Such measures imply at least Level 5 efficacy data or models. But such studies, as discussed above, are difficult to complete.

Phelps and Mushlin[48] have proposed a linkage between Level 5 conclusions and Level 2 data collection. Their model presumes that there is a societal threshold price, expressed in dollars per QALY, above which society is unwilling to invest resources, and below which the investment is not questioned. This is an oversimplification, but in general there seems to be a threshold above which at least questions are raised, and below which there is less overt concern. For the sake of argument this threshold could be placed around $50,000 per QALY. Phelps and Mushlin consider a new imaging test (such as magnetic resonance imaging) that may replace an existing imaging procedure (such as CT) for a particular medical diagnosis and management problem. Given the ROC parameters for the existing procedure are known, i.e., that a Level 2 study has been done for the existing procedure, they compute a "challenge region" in the ROC diagram into which the ROC curve for the new procedure must fall if its incremental cost per QALY is to be less than the societal threshold.

Using the Phelps and Mushlin approach, a cost–effectiveness model (really a decision analysis of whether to use the new procedure or not) addressing issues about incremental patient outcome efficacy helps to reduce the empirical task to that of conducting a Level 2 study—a task much more feasible than conducting a prospective randomized trial.

Thus, the relationship of the hierarchical model proposed for efficacy studies in diagnostic imaging is intimately related to cost–effectiveness analysis. Use of the model provides a language to discuss different-appearing, but related, research designs within one framework.

Conclusion

Efficacy studies provide a way of understanding—and thus comparing—the usefulness of diagnostic imaging examinations and procedures. Efficacy of diagnostic imaging is complicated by the fact that imaging is one step in a larger process and there are many points in this process to measure the information transmitted or the effects of the information on the process. Each measure has some legitimacy in being called an efficacy measure and each investigation using them can be called an efficacy study. Only by examining the full continuum do we begin to see the interrelationships and the limitations as well as strengths of each study design.

A good deal of this model for diagnostic imaging is directly applicable to any mode of seeking and using diagnostic information. We have drawn attention especially to diagnostic imaging as it is this field in which Dr. Lusted (whom this issue of *Medical Decision Making* honors) has made his special seminal contributions. And, in particular, a good deal of the early impetus for efficacy study and the important application of ROC analysis in diagnostic imaging can be traced directly to Dr. Lusted's pioneering work.

The model described in this paper was formulated in discussions of Scientific Committee Number 69 of the National Council on Radiation Protection and Measurements. Portions of the text draw on draft materials produced by that working group. The group was chaired by one of the authors (JRT) and the other (DGF) was a member. The authors gratefully acknowledge the contributions of the other members, Robert A. Goepp, DDS, Lee B. Lusted, MD, Keith I. Marton, MD, Barbara J. McNeil, MD, PhD, and Milton C. Weinstein, PhD, who were involved in the initial model development meeting. Others subsequently involved in helping to elaborate aspects of the model were Kunio Doi, PhD, Charles E. Metz, PhD, Harvey Rudolph, PhD, Alvin I. Mushlin, MD, and Charles E. Phelps, PhD.

Downloaded from http://mdm.sagepub.com at ALBERTA UNIVERSITY on March 14, 2008
© 1991 Society for Medical Decision Making. All rights reserved. Not for commercial use or unauthorized distribution.

References

1. Relman AS. Assessment and accountability. The third revolution in medical care. N Engl J Med. 1988;319:1220–2.
2. National Council on Radiation Protection and Measurements. Ionizing radiation exposure of the population of the United States, NCRP Report No. 93, National Council on Radiation Protection and Measurements, Bethesda, MD, 1987.
3. National Council on Radiation Protection and Measurements. Exposure of the U.S. population from diagnostic medical radiation, NCRP Report No. 100, National Council on Radiation Protection and Measurements, Bethesda, MD, 1989.
4. National Academy of Sciences/National Research Council. The effects on population of exposure to low levels of ionizing radiation. Committee on Biological Effects of Ionizing Radiation, National Academy Press, Washington, DC, 1980.
5. Bunge RE, Herman CL. Usage of diagnostic imaging procedures: a nationwide hospital study. Radiology. 1987;163:569–73.
6. McNeil BJ, Tigansky D, Wennberg JE. The use of medical radiographs: extent of variation and associated active bone marrow doses. Radiology. 1985;156:51–6.
7. Lusted LB. Introduction to medical decision making. Springfield, IL, Charles C Thomas, 1968.
8. Yerushalmy J. Reliability of chest radiography in diagnosis of pulmonary lesions. Am J Surg. 1955;89:231–40.
9. Ledley RS, Lusted LB. Reasoning foundations of medical diagnosis. Science. 1959;130:9–21.
10. Beck JR, Pyle KI, Lusted LB. A citation analysis of the field of medical decision making 1959–1982. Med Decis Making. 1984;4:449–68.
11. Cochrane AL. Effectiveness and efficiency. Random reflections on health services. London: The Nuffield Provincial Hospital Trust, 1972.
12. Thornbury JR, Fryback DG, Edwards W. Likelihood ratios as a measure of diagnostic usefulness of excretory urogram information. Radiology. 1975;141:561–5.
13. Fineberg HV. Evaluation of computed tomography: achievement and challenge. Am J Roentgenol. 1978;131:1–4.
14. Lusted LB, Roberts HV, Edwards W, et al. Efficacy of diagnostic x-ray procedures. Reston, VA: American College of Radiology, 1980.
15. Schwartz JS, Ball JR, Moser RH. Safety, efficacy, and effectiveness of clinical practices: a new initiative [editorial]. Ann Intern Med. 1982;96:246–7.
16. Brook RH, Lohr K. Efficiency, effectiveness, variations and quality: boundary crossing research. Med Care. 1985;23:710–22.
17. Buyatt G, Drummond M, Feany D, et al. Guidelines for the clinical and economic evaluation of health care technologies. Soc Sci Med. 1986;22:393–408.
18. McNeil BJ, Adelstein SJ. Determining the value of diagnostic and screening tests. J Nucl Med. 1976;17:439–48.
19. Fryback DG. A conceptual model for output measures in cost–effectiveness evaluation of diagnostic imaging. J Neuroradiol. 1983;10:94–6.
20. Kent DL, Larson EB. Magnetic resonance imaging of the brain and spine. Is clinical efficacy established after the first decade? Ann Intern Med. 1988;108:402–24;474–6.
21. The PIOPED Investigators. Value of the ventilation/perfusion scan in acute pulmonary embolism. Results of the Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED). JAMA. 1990;263:2753–9.
22. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143:29–36.
23. Swets JA. Measuring the accuracy of diagnostic systems. Science. 1988;240:1285–93.
24. Lusted LB. Logical analysis in roentgen diagnosis. Radiology. 1960;74:178.
25. Lusted LB. Receiver operating characteristic analysis and its significance in interpretation of radiologic images. In Potchen EJ, ed. Current concepts in radiology. Vol. 2, pp. 117–130. St. Louis: C. V. Mosby, 1975.
26. Hanley JA, McNeil BJ. Statistical approaches to the analysis of receiver operating characteristic (ROC) curves. Med Decis Making. 1984;4:137–50.
27. England WL. An exponential model used for optimal threshold selection on ROC curves. Med Decis Making. 1988;8:120–31.
28. Angelos Tosteson AN, Begg CB. A general regression methodology for ROC curve estimation. Med Decis Making. 1988;8:204–15.
29. McClish, DK. Analyzing a portion of the ROC curve. Med Decis Making. 1989;9:190–5.
30. Gur D, King JL, Rockette HE, et al. Practical issued of experimental ROC analysis. Selection of controls. Invest Radiol. 1990;25:583–6.
31. Henkleman RM, Kay I, Bronskill J. Receiver operator characteristic (ROC) analysis without truth. Med Decis Making. 1990;10:24–9.
32. Griner PF, Mayewski RJ, Mushlin AI, Greenland P. Selection and interpretation of diagnostic tests and procedures: principles and application. Ann Intern Med. 1981;94(4 Part 2):557–92.
33. Poses RM, Cebul RD, Centor RM. Evaluating physicians' probabilistic judgments. Med Decis Making. 1988;8:233–40.
34. Food and Drug Administration. Assessment and modification of clinical utility in diagnostic radiology: the oral cholecystogram and the upper gastrointestinal examinations. FDA Publication 85-8250, Government Printing Office, Washington, DC, 1985.
35. Sane SM, Worsing, RA, Wiens CW, Sharma RK. The value of preoperative chest x-ray examination in children. Pediatrics. 1977;60:669–72.
36. Farnsworth PB, Steiner E, Klein RM, San Filippo JT. The value of routine preoperative chest roentgenograms in infants and children. JAMA. 1980;244:582–3.
37. Wittenberg, J, Fineberg HV, Black IB, et al. Clinical efficacy of computed body tomography. Am J Roentgenol. 1978;131:5–14.
38. Wittenberg J, Fineberg HV, Ferrucci JT, Simone FJ. Clinical efficacy of computed body tomography II. Am J Roentgenol. 1980;134:1111–20.
39. Mushlin AI, Thornbury JR. Intravenous pyelography: the case against its routine use. Ann Intern Med 1989;111:58–70;83–4.
40. Marton KI, Larkin J, Ward J, et al. The role of the upper gastrointestinal series in the evaluation of abdominal pain. Clin Res. 1981;29:322A.
41. McNeil BJ, Varady PD, Burrows BA et al. Measures of clinical efficacy: cost–effectiveness calculations in the diagnosis and treatment of hypertensive renovascular disease." N Engl J Med. 1975;293:216–21.
42. McNeil BJ, Collins JJ, Adelstein SJ. Rationale for seeking occult metastases in patients with bronchial carcinoma. Surg Gynecol Obstet. 1977;144:389–93.
43. Simon DG, Lubin MF. Cost–effectiveness of computerized tomography and magnetic resonance imaging in dementia. Med Decis Making. 1986;5:335–54.
44. Joint statement on mammography guidelines. Am Coll Radiol Bull. July 1989;45(7):4.
45. Hall FM. Screening mammography—potential problems on the horizon. N Engl J Med. 1986;314:53–5.
46. Eddy DM. Screening for breast cancer. Ann Intern Med. 1989; 111:389–99;858–9.
47. Chang P, Friedland GW. The role of imaging in screening for prostate cancer. A decision analysis perspective. Invest Radiol. 1990;25:591–5.
48. Phelps CE, Mushlin AI. Focusing technology assessment. Med Decis Making. 1988;8:279–89.

Downloaded from http://mdm.sagepub.com at ALBERTA UNIVERSITY on March 14, 2008
© 1991 Society for Medical Decision Making. All rights reserved. Not for commercial use or unauthorized distribution.