Decision Analysis in Nuclear Medicine

Sanjiv Sam Gambhir

The Crump Institute for Biological Imaging and Department of Molecular and Medical Pharmacology, The Division of Nuclear Medicine and Department of Biomathematics, UCLA School of Medicine, Los Angeles, California

This review focuses primarily on the methodology involved in properly reviewing the literature for performing a meta-analysis and on methods for performing a formal decision analysis using decision trees. Issues related to performing a detailed meta-analysis with consideration of particular issues, including publication bias, verification bias and patient spectrum, are addressed. The importance of collecting conventional measures of test performance (e.g., sensitivity and specificity) and of changes in patient management to model the cost-effectiveness of a management algorithm is detailed. With greater utilization of the techniques discussed in this review, nuclear medicine researchers should be well prepared to compete for the limited resources available in the current health care environment. Furthermore, nuclear medicine physicians will be better prepared to best serve their patients by using only those studies with a proven role in improving patient management.

Key Words: decision modeling; meta-analysis; cost-effectiveness

J Nucl Med 1999; 40:1570–1581

Multiple federal agencies, including the Food and Drug Administration (FDA), the Health Care Financing Administration (HCFA) and the Agency for Health Care Policy and Research (AHCPR), are interested in decision models for patient management. These agencies are faced with the difficult task of resource allocation during a time when the number of technologies and available methods for patient diagnostics and management are growing rapidly. How can these agencies decide which approaches to reimburse and which to deny? How can this process of resource allocation be kept objective and unbiased by political pressures? Answers to these questions currently remain unresolved, but statistical decision analysis in the form of decision models may help to provide objective measures of prioritizing various competing approaches. Decision models should not be the only consideration for making final decisions, but should aid in the process of both medical and resource allocation decision making. Because reimbursement for newly emerging technologies and radiopharmaceuticals is critically dependent on federal and private agency evalua-

tions that look for objective data in their review processes, it is imperative that nuclear medicine provide rigorous evidence-based decision models for various existing and emerging diagnostic and management procedures.

HEALTH CARE ECONOMICS

Health care economics is often referred to as pharmacoeconomics, because it is primarily the pharmaceutical companies that have the economic incentive and resources to fund outcomes research for their products. There are currently no private companies with financial incentives to fund outcomes research for contrasting management algorithms involving different diagnostic technologies. These companies may fund research in determining the accuracy of the diagnostic technologies, but rarely would they fund patient outcome issues for these same technologies. Health care spending as a percentage of gross domestic product (GDP) continues to rapidly increase with current estimates at ~18% and predicted estimates by the year 2030 of ~30% (*1*). This is in large part the result of the increase in the elderly population in the United States. Health care economics is becoming increasingly important because it is unlikely that society will tolerate 30% of GDP for medical care. To curtail spending, medical resources will have to be more tightly partitioned across many competing methodologies (both diagnostic and therapeutic).

TECHNOLOGY ASSESSMENT

Technology assessment includes a variety of analyses aimed at answering questions regarding the usefulness of a given technology for health care delivery. Original research studies, including clinical trials, can be used to answer questions as to the effect of a given technology on relevant outcome measures. Alternatively, secondary analyses can be performed to understand the existing literature through formal techniques, including meta-analysis. Formal decision analysis modeling can also be used for technology assessment.

Technology assessment is performed by many different groups, including clinical researchers and dedicated technology assessment units funded by the AHCPR. The private sector also has dedicated technology assessment groups, including Blue Cross and Blue Shield Association Technology Evaluation Center, ECRI, University HealthSystem Consortium and the Rand Corporation. The Technology

Evaluation Center (TEC) of the Blue Cross and Blue Shield Association was organized to formalize the process of scrutinizing new technologies. This group helps to differentiate those technologies with evidence for improving patient health outcomes from those for which the data are still too preliminary (2). The TEC criteria are listed in Table 1. These criteria form a basis on which to evaluate all emerging technologies. The technology must have a net benefit for health outcomes that generalizes beyond the research setting. It is important to note that the TEC criteria are focused on medical effectiveness without specifically looking at cost issues. A technology that meets the TEC criteria will not necessarily be reimbursed. Each organization may look at TEC and other criteria to reach decisions related to coverage.

Because TEC and other similar criteria involve evidence for changes in health outcomes, nuclear medicine studies must show not only accuracy and management change data but also outcomes data. It is not easy to demonstrate changes in outcome when analyzing an emerging diagnostic technology. Patients would have to be followed over extended periods of time (as in randomized clinical trials involving new therapeutics) to determine net health outcome changes. However, it may be possible to use small clinical trials to determine accuracy data along with patient management changes and then infer impact on health outcomes using mathematical decision models. When results from multiple smaller clinical trials are combined for the purpose of a technology assessment, it is important to include emerging unpublished data, which may lag behind the sparse published literature. Nuclear medicine clinical trials could benefit significantly if study design issues consider the TEC criteria.

META-ANALYSIS

Clinical Research

The term meta-analysis was first used in 1976 (3) to describe the efficacy of psychotherapy. Meta-analysis grew out of the psychology and education literature and is relatively common in the medical literature, but still infrequently found in nuclear medicine literature. The nuclear medicine literature focuses more frequently on literature reviews without a formal meta-analysis. As the number of

published and unpublished trials continues to increase, methods for systematic synthesis of research results are needed. High-quality meta-analysis is needed especially for providing data for decision models. Meta-analysis also can be used independently to provide overall estimates of accuracy (e.g., sensitivity and specificity) and proportion of patient management changes. A high-quality meta-analysis is needed for providing data for decision models.

Meta-analysis is the technique of combining study results to strengthen conclusions about the individual studies when taken as a whole. The approach, when properly applied, would allow investigators to determine if a given set of imaging studies support the diagnostic efficacy of the modality under study. The end result of the meta-analysis has both qualitative and quantitative elements. Numerical results of the sensitivity and specificity measurements, sample sizes of the individual studies and issues related to study quality, study design and extent of bias need to be considered. Furthermore, as more studies become available, the meta-analysis can be updated easily to remain current with the latest literature results. For example, one could take all the small clinical studies performed by various groups on the role of FDG PET in staging recurrent melanoma and perform a meta-analysis to assess the overall sensitivity and specificity supported by these studies. If the studies also addressed the important issues related to patient management changes, one could potentially determine the overall pooled probability of management changes.

Meta-analysis should be viewed as a more formalized approach to literature review and as an approach that should supplement, not compete with, a standard literature review. Although no standardized methods exist for performing and reporting a meta-analysis, several reviews have covered the important issues to be considered (4–6). The first, most critical factor before starting a meta-analysis is to define the scope and goals of the project. Are two technologies to be compared? What is the clinical question and the population of interest? These and other issues are critical to define before launching a literature review for meta-analysis.

Literature Review

Literature searching can begin with searches of computerized databases, such as MEDLINE, using key words, subject items, etc. This initial search needs to be supplemented with additional searches of other databases, reviews of abstracts from meeting publications and discussions with researchers actively involved in the field. Consultation with professional library services specialized in general literature searches may also be useful. The search methods need to be reported explicitly in the final meta-analysis so that future investigators can repeat the search for updating previous results. A detailed reporting of methods used to obtain the literature items is critical in helping to determine the completeness of the meta-analysis.

"Publication bias" refers to possible bias introduced by differences in published literature compared with unpublished studies. Although one might initially think that it is

TABLE 1
Blue Cross and Blue Shield Association Technology Evaluation Center Criteria

The technology must have final approval from the appropriate government regulatory bodies.
The scientific evidence must permit conclusions concerning the effect of the technology on health outcomes.
The technology must improve the net health outcomes.
The technology must be as beneficial as any established alternatives.
The improvement must be attainable outside the investigational setting.

inappropriate to included unpublished studies, this is an active area of debate. Investigations that do not show significant efficacy for a given diagnostic test may be less likely to be published than studies that do show efficacy. However, the bias is not always necessarily in this direction (*4*). Unpublished studies can be located through discussions with senior investigators, by following up on abstracts or through resources at the National Institutes of Health.

"Confirmatory bias" refers to a bias introduced by reviewers of papers who tend to believe data that support their views and discredit data that does not. Newer studies or "unpopular" data tend to be underreported in the published literature. An initial meta-analysis should be performed with only peer-reviewed published data, and then the results tested by performing a sensitivity analysis (a technique detailed later) if unpublished data are also included.

Once all studies to be potentially included in an initial meta-analysis are identified, inclusion and exclusion criteria can be applied. There are no universal inclusion and exclusion criteria that can be applied to all meta-analyses. The criteria depend on what objectives are sought in performing the meta-analysis. If the objectives are to determine management changes, then only studies that include such data can be included. One can always perform a meta-analysis with relatively broad inclusion criteria and then perform a sensitivity analysis to determine how the meta-analysis changes if the inclusion criteria are made more strict.

Next, the literature should be reviewed to extract summary data. This process should be performed independently by at least two investigators using a predetermined standardized form. Each investigator should be familiar with the area of investigation being explored in the literature. Raw numbers should be recorded as opposed to summary measures when available in the reported results. Relevant data are often missing in published studies, and, if possible, investigators should be encouraged to contact the authors to resolve specifics and inconsistencies.

Assessment of study quality is the next step in performing a rigorous meta-analysis. This is perhaps the most subjective portion of a meta-analysis. Methods to make this process more objective include reviewing each study without knowledge of the authors and institutions involved. Quality scores using a prespecified range can be used to provide an estimate of the overall summary of the assessed quality of each study by each reviewer. Formal methods for performing quality assessment have been addressed in the literature and are reviewed elsewhere (*4*). Several specific objective measures can be used to address the quality of nuclear medicine studies using the following criteria:

1. The population being studied should be well described (e.g., age, sex). Inclusion and exclusion criteria for the patients being studied should be explicit. The referral pattern for the patients and issues of co-morbidity for the patients should be made clear.

2. The details of the instrumentation, acquisition and image reconstruction protocols should be clearly described. Patient preparation issues should also be adequately described.

3. The reference standard(s) against which the imaging study is being compared (e.g., biopsy) needs to be specified. The limitations of the reference standard in serving as a true "gold standard" need to be made clear. This becomes particularly important for true-negative patients, in whom follow-up is usually the only indirect method of determining the "truth."

4. The imaging test and reference standard should be read independently of each other. If multiple imaging tests are involved, they should be read independently of each other. It may also be necessary to read one or more of the imaging studies in conjunction with each other (e.g., a CT and FDG PET study) if, in the final diagnostic algorithm, this is what will happen practically. The readers performing analysis of the image study should be optimally chosen to be outside of the group performing the study to minimize reader or interpretation bias.

5. If two ore more imaging studies are being compared, they should be performed on the same patients, or patients should be randomly allocated across all imaging studies being compared.

6. Confirmation of the reference standard preferably should be done in all patients. If this is not possible, a random sample of the patients should be confirmed. If all patients cannot have confirmation performed, then statistical adjustments for sampling fractions need to be performed. Verification bias occurs because often more of the test-positive patients are verified by the reference standard, but not all test-negative patients are. Even if the test-negative patients are randomly selected, this can lead to verification bias unless appropriate adjustments are made (*7*).

7. Appropriate reporting of the raw numbers to allow future pooling of the data with other literature should be provided. Final summary measures (e.g., sensitivity and specificity) should be provided along with confidence intervals. The reporting of confidence intervals for sensitivity and specificity (*4*) is rarely done with nuclear medicine literature and is critical for understanding estimates of the certainty of the point estimates.

These criteria are by no means inclusive of all possibilities but serve to form a list of the key issues helping researchers to decide on the quality of the study. Additional criteria usually will be needed, depending on the exact reason(s) for which meta-analysis is being performed.

Statistical Issues

Several statistical issues become important as one seeks to combine data across studies. These include issues of how and whether data across studies can be combined, whether variations in study results can be explained by differences in

study characteristics and what are the best estimates of the summary measures and confidence intervals for the final summary estimates. Combining sensitivity and specificity data for diagnostic studies has been addressed (6). The overall sensitivity and specificity across several studies are not always combined appropriately by performing a simple pooled or weighted averaging adjusted for sample size, because each study may have used different explicit or implicit thresholds. In general, estimating the sensitivity and specificity separately underestimates the sensitivity and specificity. Methods to fit data to a summary receiver operating characteristic curve (SROC) have been developed and are detailed elsewhere (8). In cases in which there are very few studies (3–5), it may not be possible to get a best fit to an SROC. In these cases, there will be no choice but to use a pooled or weighted estimate for the sensitivity and specificity. Logistic modeling procedures (9,10) can sometimes be used when there are no threshold differences between primary procedures. Additional methods are available for dealing with tests with nonbinary results and for additional complexities introduced by the study design (6).

Applicability

A well-performed meta-analysis that is based on well-designed primary studies still may not be useful for a given situation. A nuclear medicine physician must weigh how applicable the results of the meta-analysis are for his or her specific population of patients. Similarly, to use a meta-analysis for a decision model, one must be careful that the meta-analysis and decision model reflect the same clinical population. Patient characteristics such as age, sex, referral pattern, co-morbid conditions and other imaging and nonimaging tests performed before the imaging study of interest might all be significant factors affecting the study results. One very important factor that must be weighed is the patient spectrum. The extent of the disease in the "diseased" group and the occurrence of other medical conditions in the "nondiseased" group play an important role in generalizability. The patient spectrum is very likely to differ across practices. This is one of the reasons to be careful when applying results based on tertiary care centers to those of the primary care setting or vice versa. For nuclear medicine procedures, there can also be significant variability in

technical issues related to performing and interpreting the imaging study of interest.

DECISION MODELING: THE BASICS

Decision modeling is a general methodology for objectively choosing between two or more strategies. We all perform decision modeling in some form or another throughout our professional and personal lives. Consider an example where an individual debates whether to put $10,000 in a mutual fund that invests in the stock market, in an individual company stock or in a fixed interest-bearing account for a period of 1 y. Most individuals would consider the primary variables involved (e.g., mutual fund track history, mutual fund top holdings, mutual fund management fees, stock market average returns during the last decade, company profits or earnings, fixed interest rate for interest-bearing account) to arrive at an educated decision. Very few individuals would structure their decision as a formal mathematical model to optimize their investment. A collection of the mathematical modeling approaches to optimize a given outcome lies in the domain of decision modeling. Many mathematical modeling approaches for performing a decision analysis are available, including Markov models and decision trees. Markov models are used to model series of events with a finite number of outcomes. The outcomes usually represent health stages (e.g, healthy, newly diagnosed, dead, etc.). Transition from one health state to another is modeled with conditional probabilities placed in a large matrix. Markov models are especially useful when time needs to be explicitly modeled, as in screening programs. For further details of Markov-type approaches, the reader is referred to Sonnenberg and Beck (11). For the lay person, it is easier to relate to and understand the mathematics for decision trees.

As an example of a decision model, consider the decision tree for investing money by one of three different strategies, as shown in Figure 1. The three primary strategies are (1) invest in a fixed interest-bearing account, (2) invest in an individual stock in the U.S. stock market or (3) invest in a mutual fund with a long track record over the last 10 y. The key variables, baseline and range of these variables are listed in Table 2. The square node represents a decision node. From
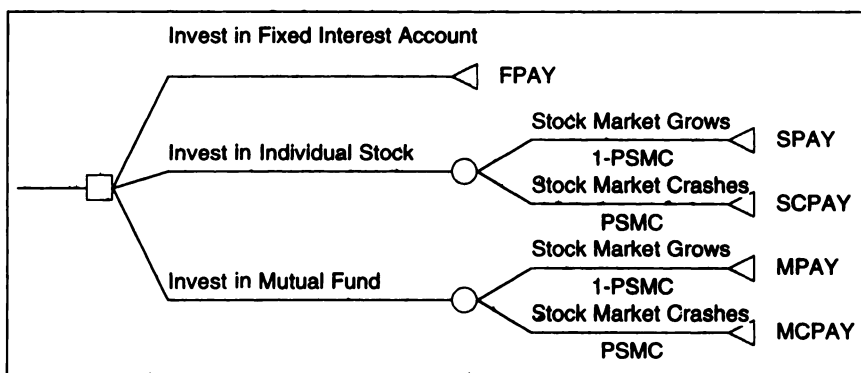

FIGURE 1. Decision tree model reflecting three possible financial investment strategies for investing $10,000 for period of 1 y. Square node represents decision node. Circular nodes are chance events, and triangular nodes are endpoints or payoff nodes. Three investment strategies include investing in fixed interest account, investing in individual stock and strategy that invests in mutual fund that invests in various stocks. Variable PSMC represents probability of stock market crashing during 1-y period under consideration. Various payoffs (e.g., FPAY, SPAY) represent final values of initial investment period after 1-y period.

TABLE 2
Baseline Variables for Investment Decision Tree Shown in Figure 1

Definition	Variable	Baseline	Range
Initial investment amount	I	$10,000	—
Fixed interest account rate	FI	5%	4%–7%
Mutual fund estimated return	MI	15%	10%–20%
Individual stock estimated return	SI	30%	5%–40%
Probability of stock market crash	PSMC	0.15	0.02–0.50
Individual stock return if market crashes	ISRMC	−30%	−50%–+10%
Mutual fund return if market crashes	MFRMC	−10%	−25%–+5%

the decision node originate three distinct strategies. The individual stock and mutual fund strategies have a circular node, which represents a chance event. This chance event represents the possibility that the stock market may crash. Because both the mutual fund return (based on the stock market) and the stock market return are dependent on this chance event, both strategies must model this event. At the end of the decision trees are triangular nodes representing endpoints or payoffs. This is where the total investment value at the end of the year is calculated. For example, for the mutual fund strategy in the case that the stock market does not crash, the payoff is equal to the initial investment × (1 + mutual fund estimated return percent/100). In the baseline case this is $10,000 × (1 + 0.15) = $11,500.

The model shown is an obvious oversimplification but serves to illustrate the process of decision modeling. The decision to be made is which is the best investment strategy, given our understanding of all the variables that will effect the outcome at the current time. The current model assumes that we wish to invest $10,000 by choosing only one of three available strategies. The fixed interest strategy will have a fixed return of 5% over a 1-y period. The mutual fund, based on our analysis of past performance and the expected conditions for this year's stock market, is expected to give a return of 15%. However, if the stock market crashes, then we estimate a loss of 10% of our original investment. The individual stock that we are considering is expected to return

30%, unless the stock market crashes, in which case we expect to lose 30%. Furthermore, the probability of the stock market crashing during the next 12 mo is estimated at 15%. For the purposes of simplification, we have not specified in this example how the baseline and range estimates of all the variables were arrived at, but let us assume that rigorous analysis (e.g., a meta-analysis) of the financial markets, stock and mutual fund data allowed us to arrive at the baseline values and ranges of this model. These analyses and the variables included could be explicitly modeled within the decision tree model or outside the decision models, as has been done in this case. The key is that a formal analysis of the supporting literature must be performed before arriving at a structure of a decision model. This analysis will also lead to baseline estimates of the mean and ranges of underlying variables. In the case of financial modeling, this may come from historical data, financial articles, and other sources. In medical decision modeling, these data usually come from a meta-analysis of the medical literature.

Analysis of the decision model is clear for the fixed investment strategy. Here, the outcome 1 y from now will be an investment worth $10,500 (assuming the 5% fixed return). The outcomes for the other strategies are not possible to predict exactly because they depend on certain events with probabilistic outcomes. Analyzing the decision tree model (a process called "rollback") for baseline estimates of all variables leads to the values shown in Figure 2. As expected, the FPAY value is $10,500. The other payoffs (SPAY, SCPAY, MPAY and MCPAY) are as shown. The process of rollback while trying to maximize returns leads to the choice of the individual stock strategy as the most likely to be the winning strategy with a mean expected value of the investment to be $12,100. The mutual fund strategy gives a mean expected value of $11,125, and the fixed investment strategy gives $10,500. The $12,100 is calculated through the formula 0.85 × $13,000 + 0.15 × $7,000. It is important to realize that the final return on our investment, if we choose the individual stock or mutual fund strategies, will probably not be $12,100 and $11,125, respectively. These values represent the expected value if $10,000 could be invested under identical conditions during thousands of separate such scenarios. Sometimes a strategy
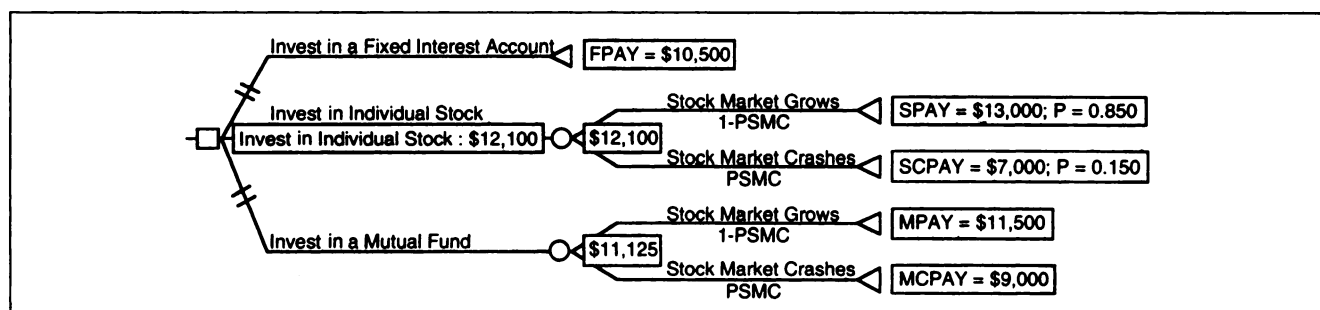

FIGURE 2. Rollback of decision tree shown in Figure 1 with baseline variables as described in Table 2. Winning strategy under assumptions of model is to invest in individual stock with mean investment value of $12,100. Next best strategy is mutual fund strategy with mean investment value of $11,125. Finally, losing strategy is fixed interest strategy with mean investment value of $10,500.

would lead to more than the expected value, and other times less, but on average we would get the value calculated as the expected value. As such, the rollback values are general statistical estimates and do not represent any realization for a given year. The expected values for each strategy, however, are the statistical estimates for guiding the investment choice. Of course, if no chance nodes are involved, as in the fixed interest strategy, then the expected value is fixed and is equal to the exact value that one would obtain. Based on the results of Figure 2 one might choose the investment in an individual stock in hopes of maximizing the return. The model, of course, can be rolled back again and again under different assumptions of the underlying variables to further add confidence to the decision. Additional outcome variables other than total investment value can also be modeled. For example, one might wish to model the level of investor anxiety with each strategy and wish to maximize the ratio of investment return to investor anxiety. Anxiety would probably be the lowest for the fixed interest strategy. The methods for objectively quantitating anxiety levels would have to be well understood by the individual performing the modeling.

Readers of decision tree results are often curious about the statistical validity of the results obtained. For example, are the values of $11,125 and $12,100 for the expected values of the mutual fund and individual stock strategies significantly different? The way to answer this question is not by performing a statistical test, but to perform sensitivity analysis and determine whether the gap can narrow under realistic changes in the baseline variables. P values (for showing statistical significance) are not calculated for decision trees, and this sometimes causes confusion for new readers of these types of model results. Applying decision trees for medical management algorithms can be accomplished by direct extension of the techniques discussed for this investment example. In the medical management scenario, one usually wishes to minimize costs and maximize health outcomes as detailed next.

DECISION ANALYSIS

Cost-Minimization

In this type of analysis, the objective is to compare strategies to determine which strategy minimizes costs. This type of analysis is fairly limited, except in cases in which the strategies to be compared are thought to have near-identical health outcomes. Obviously, if one strictly wished to minimize health care costs, one would simply use a strategy that did absolutely nothing for the patient. This would clearly minimize costs but would have devastating effects on patient outcome. Even if an analysis only includes cost issues, it may be useful in defining the dollar savings and the number of procedures avoided.

Cost-Effectiveness

Cost-effectiveness analysis (CEA) is a collection of problem-solving methods, in which a limited amount of resources must be used to assess the efficiency with which various medical management algorithms produce health outputs. The results of applying these methods are usually reported in cost-effectiveness ratios that reflect summary measures of the costs of achieving a unit of health effect (e.g., the cost per year of life gained). CEA is an aid to the decision-making process but not a cure-all for the process of decision making. Many additional factors should be considered in addition to considering the results obtained from a formal CEA. The incremental cost-effectiveness ratio (ICER) can be used to directly compare the cost of a proposed strategy with some baseline strategy divided by the life expectancy (LE) difference between the proposed and the baseline strategy. Mathematically, ICER ($COST_{new}$ − $COST_{baseline}$) ÷ (LE_{new} − $LE_{baseline}$). Ideally, it would be optimal to have a new strategy that has lower costs than the baseline strategy and leads to a greater life expectancy. This would result in a negative ICER because of the negative numerator. In many cases, however, the new strategy leads to a gain in LE compared with the baseline strategy, but at a greater cost (leading to a positive ICER). In this latter case, the payer of the health care delivery must ask whether the additional cost is worth the additional benefit. Typically, ICER values of $50,000 per year of life saved are quoted as "acceptable" for a new medical management strategy (12). Alternately, ICERs can be compared to previously published ICERs of procedures and interventions already accepted in the medical community.

Different cost perspectives can be considered when performing a CEA. From the societal perspective, the cost perspective is the net burden on the gross national product (GNP). From this perspective all individuals are included, regardless of who bears the cost (e.g., HMOs, government, employers or individuals). Several components must be considered when performing a CEA from the societal perspective. These include: (a) direct health care costs, (b) direct personal costs, (c) direct nonhealth costs and (d) indirect costs. Direct health care costs include those directly related to the new medical intervention, downstream costs induced or avoided, costs of complications incurred, savings because of avoidance of morbidity, etc. Costs of treating conditions during added years of life usually are not added. Direct personal costs include costs of patient transportation, home equipment, etc. Direct nonhealth costs include costs or savings that are directly attributable to the medical intervention but are nonmedical in nature. An example would be lawsuits related to a screening strategy that missed a tumor, leading to tumor progression during the observation period. Indirect costs include productivity gains or losses by a patient directly associated with the disease process and its management, and opportunity costs, such as travel time. Not all of these four components need to be modeled for every CEA, but each needs to be considered before deciding which has a major economic impact on the particular study and, therefore, may need to be included.

Several perspectives other than the societal perspective

are reasonable approaches to performing a CEA. These include the perspectives of the government, government health care payers, managed care, hospitals and consumers. These other perspectives can be important because the societal perspective does not optimize a CEA from a specific subgroup perspective. The societal perspective does not treat any subgroup as more important than another. Only the societal perspective never counts as a gain what is any particular subgroup's loss, For CEA involving nuclear imaging studies, any one of the above perspectives would be appropriate, but primarily the societal and government health care payer perspectives have been used.

In the day-to-day practice of medicine, it is very difficult to keep the societal perspective in mind. Consider, for example, a low-yield therapeutic protocol with a 5% chance of putting a cancer patient into remission at a cost of $100,000. Now consider that the same $100,000 could be used to perform 1000 screening tests for cancer at a cost of $100 per test in which early detection is known to have a "high" probability of cure. From a societal perspective, it is difficult to justify the use of the $100,000 for one person, when 1000 people could get the screening test instead. From the individual's perspective, however, the 5% chance of remission represents a significant hope that is difficult to disregard. Nevertheless, the societal perspective is likely to do more justice to all patients within the medical practice than to each individual patient.

Also important in CEA is the pretest likelihood of disease in calculating the number of individuals who are test positive and test negative in a medical decision tree. Shown in Figure 3 are the basic probabilities involved in a medical decision model for an imaging study with a binary outcome. The post-test probability of disease can be calculated easily in both the test-positive and test-negative groups and propagated further down the decision tree for use in the next test (assuming independence of tests). Shown in Figure 4 is an example of the use of decision analysis with decision trees for the preoperative staging of non-small cell lung cancer (NSCLC) (13,14). This decision model compares the conventional strategy of using CT alone with that of using

CT and FDG PET. The patients are already diagnosed with NSCLC and are being preoperatively staged. The key is to detour patients who, because of the spread of NSCLC, would not benefit from a thoracotomy. This decision model was used to show under what conditions the CT and FDG PET strategy is cost-effective while using life expectancy for the effectiveness criteria.

Discounting

Recommendations for CEA (15–17) include discounting both costs and life expectancy. Discounting costs adjusts for the fact that the current and future value of money is not the same. The general recommendation is to use 3% above inflation per year, but values as high as 5% have been used (17). Discounting health effects is more difficult to understand, but if one thinks of an example in which the costs of two strategies are equivalent, but in one strategy the life-expectancy gain is immediate and in the other the life-expectancy gain is not realized until many years in the future, then a need for discounting makes sense. There is also some support for the discounting rate to be different for cost versus life expectancy in some cases (17,18), but for a baseline analysis they can be made equivalent. Discounting can make an important difference in the final ICER value, especially in those models involving significant amounts of time (e.g., long-term prevention programs). Discounting is not modeled in the decision model of Figure 1. If we were looking at investments over multiple years, then it might be appropriate to discount because the future value of money would not be equivalent to the current value.

Life Expectancy and Quality of Life

It is generally more difficult for investigators to relate to life expectancy and quality of life than to cost issues. Life expectancy should be viewed in statistical terms, with the understanding that values used are the mean or median values for large cohorts of individuals. If a model states that the mean life expectancy of a healthy 64-y-old white man in the United States is 15 y, this means that on average, as we consider thousands of such individuals, some will live less than 15 y, others more than 15 y, but on average the
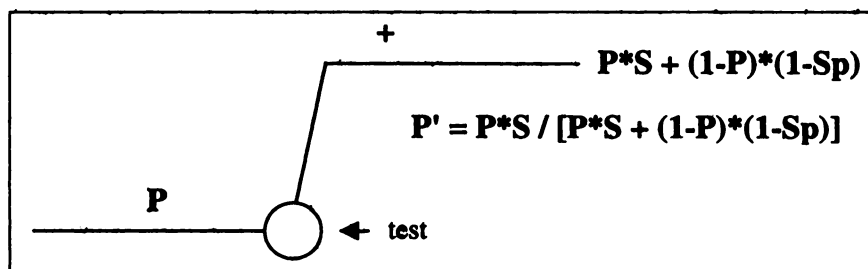

FIGURE 3. Probabilities involved in binary outcome imaging study ("test") for use in decision tree model. Imaging study is represented by chance node in decision tree. Pretest probability of disease P directly determines number of true positives and false-positives in conjunction with sensitivity (S) and specificity (Sp) of test. Probability of disease in test-positive patients is denoted by post-test probability P'. P*S = true-positive; (1 − P)*(1 − Sp) = false-positive; P' = (true-positives ÷ [true-positives + false-positives]). Similar approach can be used to determine probability of disease in test-negative patients. Post-test probabilities such as these can be propagated down decision tree.
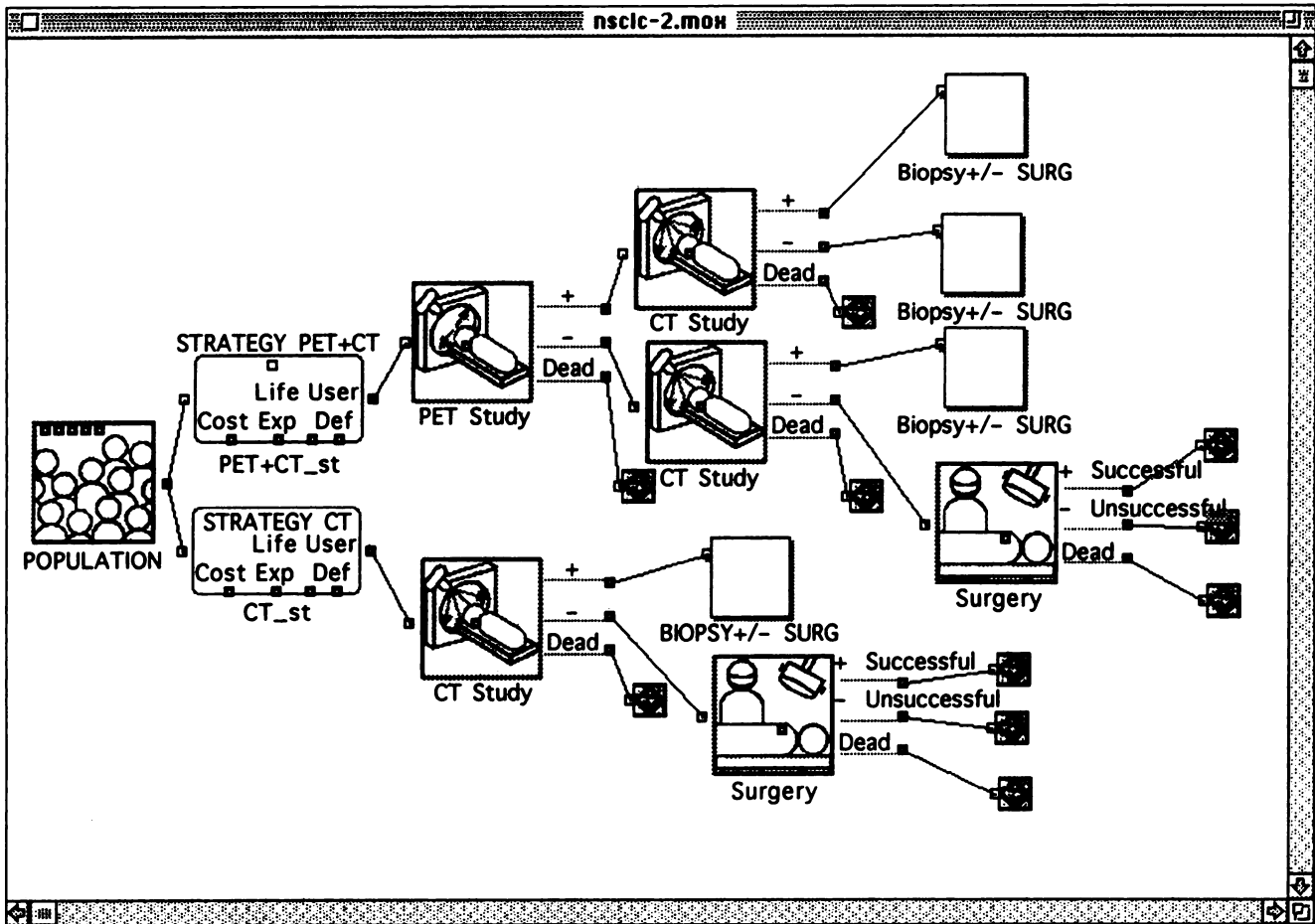
FIGURE 4. Decision tree for preoperative staging of non-small cell lung cancer using two management strategies (PET + CT_st and CT_st). Upper strategy uses both CT and PET, whereas lower strategy uses only CT. Population block describes characteristics of entering population, such as prevalence of mediastinal and contralateral lung cancer. CT and PET blocks produce several false positives and false negatives, depending on specificity and sensitivity, respectively. Biopsy ± surgery block contains details of performing biopsy and sending patient to surgery if biopsy determines no evidence of mediastinal involvement. If biopsy is positive, then patient does not go to surgery. Associated with surgery block are morbidity and mortality of thoracotomy. Small stop blocks at end of decision tree are terminal nodes.

life expectancy will be 15 y. How are investigators arriving at these mean life-expectancy values? Insurance companies that use actuarians specialize in determining the mean life expectancy of various populations and subgroups. Furthermore, because various risk factors shorten life expectancy, several methods are available for determining mean life expectancy in cases of associated disease. The declining exponential approximation of life expectancy (DEALE) is a simple and effective method for modeling life expectancy, in which an exponent term is inversely related to the mean life expectancy (19). Different exponential terms can be added to arrive at an overall mortality rate. The DEALE approximation has been used in many medical decision models and is a good initial method for dealing with many life-expectancy issues.

Although it is important to model life expectancy in a decision model, most individuals and societies understand that it is not only the quantity of life that is important but also the quality of life. The problem, however, is that we all value quality of life in different ways, and it may be difficult to

quantify objectively. Nevertheless, well-validated methods exist to measure the quality of life. Quality-adjusted life years (QALYs) are a commonly accepted method to merge the concept of life expectancy with quality of life. Methods for collecting quality preferences involve asking patients or members of the public to locate their preferences for health states on a scale of 0–1. Techniques such as time trade-off, standard gamble and category rating can be used (20). Preferences can also be measured indirectly using systems such as the Quality of Well-Being Scale and Health-Utilities Index (21). The use of QALYs in place of life expectancy in the ICER is a recommended method for CEA (15). QALYs do have some disadvantages that must be kept in mind. There is no difference in the absolute value of the QALY between an intervention that produces small benefits for many people and one that produces large benefits for very few people. QALY is a good measure that combines life expectancy and quality of life, but it cannot reflect all issues related to measuring effectiveness.

Costs

Cost issues are easier to understand conceptually in a decision model than life-expectancy issues, although what specific costs to use in a decision model are not always easy to determine. The units for U.S. health care costs are easy: U.S. dollars. Total costs must be modeled to include initial and downstream costs. These downstream costs will depend on probabilistic events that also must be appropriately modeled. Most cost data that are accessible to researchers are charges, not true costs. Charges can be nonuniformly related to true costs, making it difficult to use charges in a decision model. Furthermore, charges can vary significantly across different practices. Ideally, one would like to use an economic task force to perform true costing of a procedure, including depreciation costs of equipment, mean labor costs, etc. Practically, it is very difficult to perform rigorous costing of all components of a decision model. One method that is being used by investigators publishing in various journals is to use Medicare-reimbursed costs as a way to normalize costs across all procedures. This is appropriate when the government health care provider CEA perspective is used but may not be optimal when using the broader societal perspective. However, a CEA with the government health care provider perspective may be useful for arriving at first-order estimates before performing a CEA with a societal perspective.

Cost-Benefit Analysis

This analytical method is the same as cost-effectiveness analysis, except that the units to measure the outcome of a treatment are in dollars, not in terms of life expectancy or quality of life. This means that if a given intervention adds, on average, a few years to patient life expectancy, then this must be expressed in dollars before using it as an outcome measure in a cost-benefit analysis. The terminology is confusing, but for the most part, because health outcomes are not measured in dollars, cost-effectiveness analysis is the methodology used for decision modeling in the health care setting.

SENSITIVITY ANALYSIS

This analytic method is the single most-powerful technique available to modeling of many types. This method allows the investigator to ask "what if" questions to explore a model under various conditions. From a mathematical perspective, a model is composed of numerous variables that affect one or more outcome variables. Sensitivity analysis changes one or more variables and determines the effect this has on the outcome variable(s). For meta-analysis, a sensitivity analysis could be used to determine the effects of including articles that do not meet certain inclusion criteria or studies with a lower quality cutoff value than those used for the initial analysis. For decision analysis modeling, this approach is used easily to vary model parameters, such as the cost of a given study, to determine the effect this has on the mean overall costs per patient for a given strategy. This approach is very powerful because one can determine how sensitive the outcome(s) of a model is for a given range of a model variable.

A sensitivity analysis for the decision tree in Figure 1 with respect to the variable representing the probability of the stock market crashing (PSMC) is shown in Figure 5. As shown, the fixed interest strategy returns the same fixed amount of $10,500 independent of the probability of the stock market crashing. The other two strategies return different expected amounts based on the probability of the stock market crashing. The "threshold" value of 0.417, where the individual stock and fixed interest strategies intersect, reflects the fact that these two strategies are expected to have identical investment returns if the probability of the stock market crashing during the year is truly as
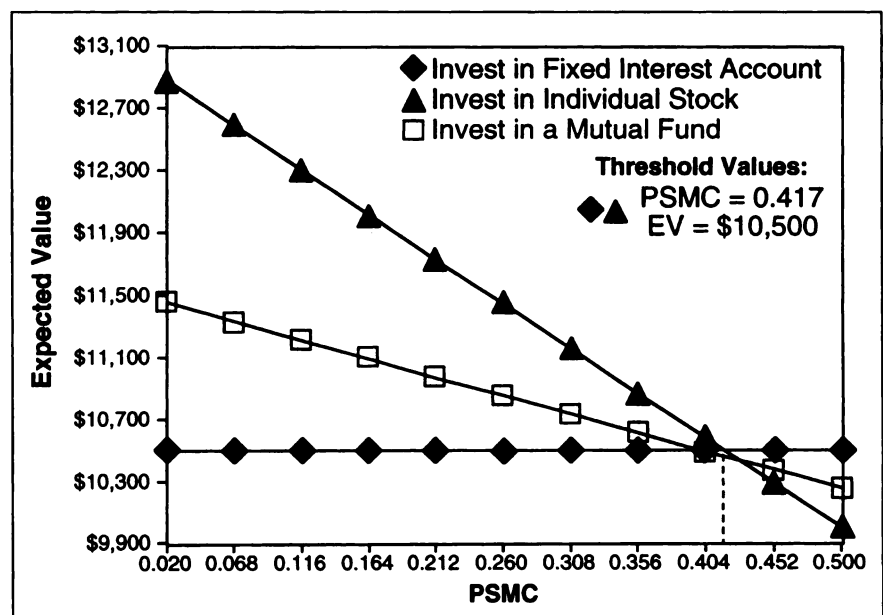

FIGURE 5. Sensitivity analysis with respect to probability of stock market crashing (PSMC) (x-axis) with total investment expected value as outcome variable (y-axis) for decision tree shown in Figure 1. The intersection of fixed investment and individual stock strategies occurs at value of PSMC = 0.417 and is referred to as threshold value. At this threshold value both strategies are expected to have equivalent investment expected values of $10,500.

high as 41.7%, assuming all other variables remain fixed at their baseline values. Thresholds for other intersection points can be determined similarly. To vary two or more variables, extensions of the same technique also can be used. An "n-way" sensitivity analysis is the most general way to explore multiple variables simultaneously. However, sensitivity analysis of more than two variables is difficult and sometimes impossible to present graphically and comprehend. Therefore, an approach that penalizes multiple variables simultaneously is sometimes used. This method penalizes several or all variables that favor a given new strategy over the baseline strategy by a fixed percentage (e.g., 10%) to determine the effects on the outcome variables of health costs or life expectancy (22).

OTHER ISSUES IN DECISION MODELING

Software Tools

Although decision analysis for a new medical problem may seem quite difficult, software tools help facilitate the process. DATA (TreeAge Software Inc., Williamstown, MA) is one of the more robust packages, designed to help construct decision models, analyze the models and perform sensitivity analyses. Convenient graphical output, including plots of sensitivity analyses, makes a decision tree easy to analyze. The Computational/Communication Sciences Laboratory at UCLA, has developed a software package (MD@) (23) that is targeted toward individuals with little mathematical expertise and facilitates the construction of simple decision models without explicitly having to deal with equations. MD@ may be particularly useful for the beginning modeler who may not want to get distracted by the underlying mathematics. Several available software packages have been reviewed (24). The cost of various decision tools is not prohibitive, with single-user licenses currently ranging from $200 to $600, with versions available for both PC and Macintosh computers.

Decision Models for Design of Future Technologies

Perhaps one of the most underappreciated uses of a decision model is the ability to predict how future technology should be designed to for a more cost-effective role. For example, a decision model for screening women with dense breast tissue could be used to answer the question "how sensitive and specific must a new scintigraphically based approach be to be cost effective?" CEA could also be used to determine the costs per patient study needed for cost-effectiveness in a technology yet to be engineered. As more well-studied decision models become available, they should be useful in studying future technology development.

Common Pitfalls

There are several common mistakes that investigators make when applying modeling procedures for decision analysis. First, one should attempt a bottom-up modeling approach to build the simplest model that explains the major features of the clinical problems; not attempt to model every possible issue at the outset. This is in direct contrast to top-down modeling, in which every possible complexity is modeled, followed by systematically pruning down various issues until a workable model is reached. A bottom-up model is usually more practical, because one can get lost in the myriad of possible clinical options in a top-down approach. The modeling process should also be considered iterative, with the model constantly being refined on the basis of previous analysis and continued greater understanding of the clinical problem. The biggest mistake at the outset is to not have a good clinical understanding of the management issues. If the clinical problem is not well understood, the decision model may be internally consistent but will not reflect the real clinical scenario.

A good understanding of the incidence/prevalence or pretest likelihood of disease for patients entering into a given decision tree is needed. This value will directly affect the number of individuals that a given imaging study calls out as positive or negative. When multiple sequential tests (imaging or other) are involved, a lack of independence between the tests needs also to be considered carefully. Issues that involve time in a decision tree (e.g., waiting during a screening strategy) need to be carefully modeled by accounting for the probabilities of various events during the waiting interval (e.g., death or progression of disease).

All costs should be carefully modeled. Downstream costs, including costs of terminal care, should be assessed to perform an accurate CEA. One good way of ensuring that one has modeled all the relevant costs is to make sure that every possible patient pathway is followed out until patient death. Even if a patient is "cured," his or her costs and other outcome variable issues need to be followed until death to properly compare the patient with those who were not cured. Some issues surrounding the inclusion of downstream costs remain controversial and are discussed elsewhere (25).

CEA analysis is not easily performed by a single investigator. A true collaboration is needed between individuals with clinical expertise and those with mathematical or statistical expertise. It is also important to understand the patient's perspective to have a full understanding of the clinical problem. A common pitfall is to launch into a decision model without understanding all the downstream issues that are not directly handled in nuclear medicine practice. It is extremely important to conduct many physician and patient interviews before developing the mathematical model of interest.

ROLE OF CLINICAL TRIALS

Well-designed randomized clinical trials are a commonly accepted method for evaluating new therapeutic interventions. These trials can be directly coupled to health outcomes, and the net impact of therapeutic interventions can be ascertained. Clinical trials for diagnostic technologies with outcome measurements as part of the study design are

performed rarely and, in my opinion, are not practical or necessary. It is difficult to construct trials in which the use of a diagnostic technology can be directly linked to the health outcomes of patients. However, it is possible to link performance of a diagnostic test to impact on patient management. This information can then be used to drive a decision model that considers the full impact of the diagnostic test on health outcomes. If clinical trials are then eventually possible, they may be used to reinforce the results of the decision model. In many cases in which full-scale clinical trials are not possible or practical, a well-constructed and -analyzed decision model may be used to predict the cost-effectiveness of a specific diagnostic technology as applied in a given management scenario.

NUCLEAR MEDICINE META-ANALYSIS AND DECISION MODELING LITERATURE

Meta-analysis and decision modeling are not new methodologies nor strangers to the field of nuclear medicine. However, it appears that decision modeling is generally overlooked by our field, with an occasional resurgence as it is applied to new relevant medical diagnostic or management algorithms. For the most part, our research studies focus on sensitivity and specificity determination. A few studies attempt to look at incremental gains, fewer yet ascertain patient management changes and there are rare studies that perform a formal CEA. Although it is not the purpose of this article to review all the meta-analysis and CEA publications relative to nuclear medicine procedures, some key examples relate specifically to meta-analysis (26–28), whereas others show its effectiveness in cost analysis (without considering effectiveness issues formally) for sestamibi scintimammography (29,30) and cardiac studies (31,32), in the role of sestamibi scintimammography for screening women with dense breast tissue (33) and in FDG PET in non-small cell lung cancer and SPN detection (13,14,22). For the most part, the nuclear medicine literature is very sparse with respect to both meta-analysis and decision analysis.

CONCLUSION

Meta-analysis and decision analysis are established techniques for modeling data for the purposes of supporting the use of specific technologies in medical management algorithms. With a better understanding and proper utilization of the techniques involved for these methods it should be possible for more researchers to apply them to nuclear medicine studies. This should lead to better acceptance of our procedures, more efficient utilization of limited resources and, most important, better patient care.

This review attempts to clarify some of the issues involved in meta-analysis and decision analysis. If we are to remain competitive in the medical community, we must plan ahead and make the types of analyses discussed in this article a routine part of our daily activities. If we are well

prepared, then we can continue to refine and defend our decision models.

REFERENCES

1. McCombs JS. Pharmacoeconomics: what is it and where is it going? Am J Hypertens. 1998;11(suppl):112S–119S.
2. Till M. Payer evaluates medical procedures. In: Mission KS, ed. IMDA Update. New York, NY: Independent Medical Distributors Association; 1997:14–22.
3. Smith ML, Glass GV. Meta-analysis of psychotherapy outcome studies. Am Psychol. 1977;32:752–760.
4. L'Abbé KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. Ann Intern Med. 1987;107:224–233.
5. Thacker SB. Meta-analysis. A quantitative approach to research integration. JAMA. 1988;259:1685–1689.
6. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. J Clin Epidemiol. 1995;48:119–132.
7. Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. Biometrics. 1983;39:207–215.
8. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat Med. 1993;12:1293–1316.
9. Albert A. On the use and computation of likelihood ratios in clinical chemistry. Clin Chem. 1982;28:1113–1119.
10. Irwig L. Modelling result-specific likelihood ratios [letter; comment]. J Clin Epidemiol. 1992;45:1335–1338.
11. Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. Med Decis Making. 1993;13:322–338.
12. Garber AM, Phelps CE. Economic foundations of cost-effectiveness analysis. J Health Econ. 1997;16:1–31.
13. Gambhir SS, Hoh CK, Phelps ME, Madar I, Maddahi J. Decision tree sensitivity analysis for cost-effectiveness of FDG-PET in the staging and management of non-small-cell lung carcinoma. J Nucl Med. 1996;37:1428–1436.
14. Scott WJ, Shepherd J, Gambhir SS. Cost-effectiveness of FDG-PET for staging non-small cell lung cancer: a decision analysis. Ann Thorac Surg. 1998;66:1876–1883.
15. Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses: Panel on Cost-Effectiveness in Health and Medicine. JAMA. 1996;276:1339–1341.
16. Siegel JE, Torrance GW, Russell LB, Luce BR, Weinstein MC, Gold MR. Guidelines for pharmacoeconomic studies: recommendations from the Panel on Cost Effectiveness in Health and Medicine. Pharmacogenetics. 1997;11:159–168.
17. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost Effectiveness in Health and Medicine. JAMA. 1996;276:1253–1258.
18. Weinstein MC. Principles of cost-effective resource allocation in health care organizations. Int J Technol Assess Health Care. 1990;6:93–103.
19. Stalpers LJ, Van Gasteren HJ, Van Daal WA. DEALE-ing with life expectancy and mortality rates. Med Decis Making. 1989;9:150–152. See also erratum in Med Decis Making. 1989;9:219.
20. Froberg DG, Kane RL. Methodology for measuring health-state preferences. III: Population and context effects. J Clin Epidemiol. 1989;42:585–592.
21. Kaplan RM, Anderson JP. A general health policy model: update and applications. Health Serv Res. 1988;23:203–235.
22. Gambhir SS, Shepherd JE, Shah BD, et al. Analytical decision model for the cost-effective management of solitary pulmonary nodules. J Clin Oncol. 1998;16:2113–2125.
23. Gambhir SS, Gupta P, Shepherd JE, et al. MD@: a physician-friendly decision analysis tool. MD Comput. 1998;15:40–48.
24. Bhargava HK, Sridhar S, Herrick C. Beyond spreadsheets: tools for building decision support systems. Computer. 1999;32:31–39.

25. Garber AM, Phelps CE. Economic foundations of cost-effectiveness analysis: working papers series. *Nat Bur Econ Res.* 1992;4164:1–34.

26. Littenberg B, Mushlin AI. Technetium bone scanning in the diagnosis of osteomyelitis: a meta-analysis of test performance. Diagnostic Technology Assessment Consortium. *J Gen Intern Med.* 1992;7:158–164.

27. Devous MD Sr, Thisted RA, Morgan GF, Leroy RF, Rowe CC. SPECT brain imaging in epilepsy: a meta-analysis. *J Nucl Med.* 1998;39:285–293.

28. Laruelle M. Imaging dopamine transmission in schizophrenia: a review and meta-analysis. *Q J Nucl Med.* 1998;42:211–221.

29. Hillner BE. Decision analysis: MIBI imaging of nonpalpable breast abnormalities. *J Nucl Med.* 1997;38:1772–1778.

30. Waxman AD. A perspective on decision analysis modeling as it relates to sestamibi imaging of nonpalpable breast abnormalities [editorial, comment]. *J Nucl Med.* 1997;38:1778–1780.

31. Patterson RE, Eng C, Horowitz SF, Gorlin R, Goldstein SR. Bayesian comparison of cost-effectiveness of different clinical approaches to diagnose coronary artery disease. *J Am Coll Cardiol.* 1984;4:278–289.

32. Patterson RE. Cost-effectiveness analysis in diagnosis of cardiac disease: overview of its rationale and method. *J Nucl Cardiol.* 1996;3:334–341.

33. Allen M, Hendi P, Bassett L, Phelps ME, Gambhir SS. A study on the cost effectiveness of sestamibi scintimammography for screening women with dense breasts for breast cancer. *Breast Cancer Res Treat.* 1999:in press.