FROM RESEARCH TO PRACTICE

Assessing the Effectiveness of Health Interventions for Cost-Effectiveness Analysis

Jeanne S. Mandelblatt, MD, MPH, Dennis G. Fryback, PhD,
Milton C. Weinstein, PhD, Louise B. Russell, PhD, Marthe R. Gold, MD, MPH,
and members of the Panel on Cost-Effectiveness in Health and Medicine

Cost-effectiveness analysis (CEA) is an analytic tool in which the costs and effects of an intervention designed to prevent, diagnose, or treat disease are calculated and compared with an alternative strategy to achieve the same goals. The results of a CEA are presented as a ratio of costs to effects, where the effects are health outcomes such as cases of disease prevented, years of life gained, or quality-adjusted life years gained, rather than monetary measures, as in cost-benefit analysis. Conducting a CEA requires a framework for portraying the cascade of events that occur as a consequence of the decision to intervene, for describing the probability that each event will occur, for accounting how long each event will last, and describing how much each event costs and is valued by the population or individuals targeted by the intervention. Mathematical models are well suited to these purposes.

The purpose of this article is to provide an overview of modeling to estimate *net* effectiveness in a CEA (the difference in effectiveness between an intervention and the alternative to which it is being compared). Many of the principles described for estimating effectiveness apply equally to determining costs in a CEA. The main difference is that health events are weighted by costs in the numerator of the cost-effectiveness ratio, while they are often weighted by preference values in the denominator. Preference values, or utilities, reflect the fact that individuals or populations with similar ability (or disability) to function may regard that level of functioning differently. When preferences are incorporated into CEAs, the results are generally expressed as costs per quality-adjusted life years.^{1,2} A discussion of measurement of costs and valuing outcomes is beyond the scope of this article; for further information on these, and other components of a CEA, the reader is referred else-

where.³⁻⁵ Following some definitions of terms, this article is organized into two sections describing the process of estimating effectiveness in a CEA: the first presents a review of the sources of event probabilities, and the second describes the use of modeling to estimate effectiveness.

DEFINITIONS

Effectiveness, which reflects the impact of an intervention of health in real practice settings, should be distin-

Received from the Department of Medicine, Lombardi Cancer Center (Cancer Clinical and Economic Outcomes Core) and the Institute for Health Care Policy and Research, Georgetown University School of Medicine, Washington, DC (JSM); Departments of Preventive Medicine and Industrial Engineering, University of Wisconsin-Madison (DGF); Department of Health Policy and Management, Harvard School of Public Health, and Department of Medicine, Harvard Medical School, Boston, Mass. (MCW); Institute for Health, Health Care Policy, and Aging Research, and Department of Economics, Rutgers University, New Brunswick, NJ (LBR); and Community Health and Social Medicine, City University of New York Medical School, NY (MRG).

The Panel on Cost-Effectiveness in Health and Medicine was convened by the Office of Disease Prevention and Health Promotion. The work presented in this paper is the responsibility of the authors, and does not reflect the views of the U.S. Public Health Service.

Dr. Mandelblatt's work was supported, in part, by grant RO1 HS08395, "Care, Costs, and Outcomes of Breast Cancer in the Elderly," from the Agency for Health Care Policy and Research, and Contract DAMD17-96-C-6069, "Breast Cancer Center Support Contract," from the Department of the Army (the "Cancer Clinical and Economic Outcomes Core" is one of the projects funded by this contract).

Dr. Gold's work was conducted while she served as a Senior Policy Advisor in the Office of Disease Prevention and Health Promotion in the U.S. Public Health Service, Department of Health and Human Services, Washington, DC.

Address correspondence and reprint requests to Dr. Mandelblatt: Georgetown University School of Medicine, Lombardi Cancer Center, Cancer Clinical and Economic Outcomes Core, 2233 Wisconsin Ave., Suite 430, Washington, DC 20007.

From Research to Practice, a Journal series, presents articles to heighten the clinician's awareness of research and methodology issues that have direct relevance to practice. If you wish to submit a manuscript for consideration for this series, please contact Cynthia D. Mulrow, MD, MSc, Associate Editor, at mulrow@uthscsa.edu, or contact the Journal of General Internal Medicine at (215) 823-4471 to receive the appropriate guidelines.

guished from two related concepts, efficacy and appropriateness. Efficacy refers to impact under ideal conditions (e.g., randomized controlled trials). Appropriateness reflects a broader range of issues considered in deciding whether an intervention should or should not be done, including acceptability, feasibility, and cost-effectiveness.⁶⁻⁸

Cost-effectiveness analysis calculates incremental effectiveness (and costs); that is, the difference in effectiveness (and costs) between the intervention of interest, and the next least effective (costly) alternative. This is distinguished from marginal effects (and costs), which refer to a production function, where, for instance there are other effects (or costs) associated with producing one additional unit of output.

To calculate *net* effectiveness, we must estimate the probabilities of all events that occur as a consequence of the intervention and alternative. Probabilities express the degree of certainty that an event will happen, on a scale from 0.0 (certainty that the event will not occur) to 1.0 (certainty that the event will occur). Probabilities found in the literature are often not in the form required for the CEA.⁹⁻¹⁴ For instance, a lifetime cumulative risk of breast cancer must be converted to an annual probability.

DATA SOURCES FOR PROBABILITY ESTIMATES

Cost-effectiveness analysis probability data can be collected as part of a research protocol (primary data), or can be abstracted or extrapolated from existing published research (secondary data). Event probability values should be selected or collected from the best designed and least biased sources that are relevant to the question and population under study using the following hierarchy (in decreasing order): well-conducted randomized, controlled trials (RCTs); observational data, including cohort, case-control, and cross-sectional studies; uncontrolled experiments; descriptive series; and expert opinion.¹⁵ Less rigorously designed studies drawing similar conclusions may be the best available source of data for a particular subpopulation or research hypothesis, in the absence of other data.

Table 1 summarizes the main advantages and disadvantages of different study designs as data sources for a CEA.¹⁶ Although well-conducted RCTs are generally accepted as the most powerful tool for assessing the efficacy of interventions, in CEA one is most interested in how an intervention performs in real-life (i.e., non-RCT) settings. Observational cohort and case-control studies can provide real-world data on the probabilities of particular outcomes associated with an intervention. Observational studies differ from RCTs in that the investigators do not have control over which persons receive the intervention, so that observational studies may be subject to unknown selection effects. Also, the use of case-control studies to evaluate the effectiveness of screening can be biased by the type of case or control group selected.¹⁷⁻²¹

For example, in evaluating the effectiveness of sigmoidoscopy screening to reduce mortality from colorectal cancer, cases should include all those dying from the dis-

ease, and controls should include both those with colorectal cancer who are still alive and those without cancer. Moreover, the definition of cancers in both the case and control groups should include those in reach of the sigmoidoscope and those beyond.²² This choice of study groups eliminates the lead time bias that would occur in situations if early-stage cases were compared with later-stage cases, where lead time bias is defined as resulting in earlier diagnosis, even though survival does not actually differ in screened and unscreened groups. Cross-sectional studies, case series, uncontrolled cohort studies, post-marketing surveillance, and disease and administrative databases may also provide data for a CEA.²³⁻²⁷

When there are insufficient data from any one source or when studies conflict, information from many types of good-quality studies can be combined to provide probability values for estimating effectiveness. The two major approaches are meta-analysis²⁸⁻³² and Bayesian methods.^{33,34} Expert opinion and consensus panels are other synthesis techniques used to estimate effectiveness. For example, the original Oregon priority list relied on educated guesses of experts who estimated the ability of particular technologies and practices to improve survival.³⁵ For further information on sources of effectiveness data, the reader is referred elsewhere.^{23,36,37}

MODELING TO ESTIMATE EFFECTIVENESS

Randomized control trials and observational studies cannot compare all relevant alternative program designs that may be of interest to the CEA analyst or policy maker. Models have the advantage of providing the user with the ability to manipulate an intervention program in ways that are not possible in real-time experiments with human subjects. Models can be used to extrapolate from existing data to different population groups, points in time, or disease end points. For instance, models enable exploration of the implications of different screening or treatment intervals. Models allow simulation of the effects and costs of ending a screening program at a given age, investigating questions such as whether to continue cervical screening past the sixth decade of life. Models allow examination of the implications of using different cutoff points for screening tests, such as the cholesterol level chosen for initiating dietary or pharmacologic intervention, or the bone mineral density level chosen for initiating treatment for osteoporosis. Models can also be of use in performing sensitivity analyses and threshold analyses to ask what the data parameters would need to be for an intervention to be considered cost-effective.³⁸

Types of Models

Models for estimating health effectiveness may be characterized along several dimensions: (1) the analytic methodology for accounting for events that occur over time, typically either a decision tree or state transition model; (2) application to cohorts longitudinally or to popu-

Table 1. Sources of Probability Data for Cost-Effectiveness Analysis

	Randomized Controlled Trials	Cohort Studies	Observational Studies and Case- Control Studies	Administrative Data	Synthesis Methods	Expert Opinion
Advantages	1. Randomization controls for known and unknown confounders 2. Outcome evaluation done without knowledge of study arm	1. Tests effectiveness 2. Potential for broad population inclusion	1. Tests effectiveness 2. Does not require long observation periods	1. May include data on large populations	1. Can address questions not previously posed 2. Increases power to detect effects 3. Improves precision estimates	1. Useful when few data exist
Disadvantages	1. Selection bias* 2. Usually tests efficacy 3. Limited periods of observation 4. Contamination effects†	1. Selection bias* 2. Can only control for known confounders 3. Requires long periods of follow-up	1. Selection bias* 2. Can only control for known confounders 3. Recall bias‡ 4. Difficulty in selection of proper control group	1. Limited types of data available 2. Incomplete data possible 3. Confidentiality issues	1. Quality depends on quality of original studies included 2. Publication bias§ 3. Subject to bias from cognitive heuristics	1. Process for combining data is left to the judgment of the experts 2. Quasi-scientific

*Selection bias: healthy subjects may volunteer to participate in research; subjects participating in research (or surviving to participate) may differ on unmeasured characteristics from those who do not participate.

†Contamination effects: the limited window of opportunity to conduct a trial or study prior to widespread introduction of an intervention into clinical practice.

‡Recall bias: subjects with disease may be more likely to recall past exposures than subjects without disease.

§Publication bias: studies finding a positive intervention effect are more likely to be published than those finding a negative effect or no effect.

||Cognitive heuristics: simplistic judgment and memory processing strategies to formulate probability assessments.

lations cross-sectionally; and (3) using deterministic or stochastic (probabilistic) calculations.

Analytic Methodology

Decision tree models represent chance events and decisions over time.³⁹⁻⁴¹ Each path through the decision tree represents one possible sequence of events, and is associated with a probability and a consequence, such as life expectancy, or quality-adjusted life expectancy. Decision analysis models have been used extensively in the medical literature, for example, to estimate gains in life expectancy from vaccines,⁴²⁻⁴⁴ and for screening elderly women for breast cancer.⁴⁵ One limitation of decision trees is that they are not well suited to representing multiple outcome events that recur over time.

State-transition models are more efficient representations of recurring events. State-transition models allocate, and subsequently reallocate, members of a population into one of several categories, or *states*. States may be de-

finied according to disease stage, treatment status, or a combination of the two. Transitions occur from one state to another at defined, recurring time intervals (usually 1 year, but sometimes 3 months or 1 month for rapidly progressive diseases) according to transition probabilities. Transition probabilities can be made dependent on population characteristics, such as age or other risk factors.

Through simulation, or mathematical calculation, the number of members of the population passing through each state at each point in time can be estimated. A special type of state-transition model, in which the transition probabilities depend only on the current state (and not, for example, on the previous states or the path by which the current state was entered), is called a Markov model.⁴⁶ State-transition models have been used to estimate outcomes in a large number of cost-effectiveness studies, including coronary heart disease prevention⁴⁷ and treatment⁴⁸; breast,⁴⁹ cervical,⁵⁰ and prostate cancer screening⁵¹; and hormone replacement therapy.⁵² Decision tree models can also be augmented to include "Markov nodes," or

branching points within the tree that lead into a Markov model.⁵³ Several computer programs, such as SMLTREE (© 1989, J. Hollenberg), DATA (© 1994, TreeAge Software, Inc.), and Decision Maker (© 1980, 1993, S.G. Pauker, F.A. Sonnenberg, and J.B. Wong, New England Medical Center, Boston, Mass.) can be used to construct such models. Other types of models, such as difference equations, have been used to assess the effectiveness of interventions targeting infectious diseases, such as AIDS prevention programs.⁵⁴

Longitudinal and Cross-Sectional Models

All models include a population or group that is relevant to the research question. Simulations then project future outcomes or “follow the patients or individuals over time.” There are two common ways these modeling approaches can be accomplished, differing in the way in which the study population is constituted at the start of the model.

The first method, known as *longitudinal modeling*, calculates expected outcomes for “typical” patients or cohorts (i.e., groups of 50-year-old men with a first myocardial infarction) and follows them longitudinally through time to evaluate health outcomes resulting from alternative interventions. This approach is often used in decision tree models or models to extend the follow-up period of an RCT from the end of the trial, typically 1 to 5 years, to death.⁵⁵ Results of such models are typically expressed as quality-adjusted life-years.

The second method is known as the *cross-sectional model*. These models record the health outcomes of a cross-section of an entire population, or a substrata of the population, and then follow each person in the population from their age at the start of the model to the end-point of the analysis (either a specified period, such as 10 years, or until death). The outcomes from alternative interventions are then summed or averaged over the population and expressed as an aggregate measure, such as quality-adjusted person-years. The model CAN*TROL and the Coronary Heart Disease Policy Model are examples of cross-sectional population models.^{47,56}

The choice of a cross-sectional or longitudinal model is determined by the problem being studied. For instance, a cross-sectional model may be used to ask public health questions about interventions that are to be applied population-wide to groups of varying ages; a longitudinal model may be used to ask questions about the long-term effects of an intervention on an age-specific group.

Deterministic and Stochastic Models

Deterministic models calculate probabilities as an average number of health events. For example, suppose that one is interested in the number of people in a cohort of 10,000 who will be dead in 10 years from a particular disease, and suppose further that one knows the annual disease-specific mortality rate is 10% and the average annual other-cause

mortality rate is 1%. The number of people who will be dead from the disease can be computed directly by multiplying the survival percentages by the expected number of survivors recursively in each of the 10 years.

In *stochastic models*, known as *discrete event simulations*, probabilities for each individual in the cohort over time are simulated using computer-generated random numbers to represent chance events. For instance, to calculate 10-year survival, simulating a 10% chance of death in a given year, the computer is directed to generate a random integer between 1 and 100, and if that integer is 10 or less, the computer program tallies the simulated person as dying in that year; otherwise, the person is deemed to have survived. The process is repeated over time for the survivors. The number of people in the cohort who are “observed” to live the full 10 years in the simulation is used as an estimate of the number that would be observed were a real study done under conditions of the simulation. The entire simulation is repeated many times, and the counts are averaged across simulation runs; as the number of runs grows large, these averages approach the values that would be computed by deterministic calculations. This type of discrete event simulation is also known as Monte Carlo simulation (e.g., the MISCAN simulation program⁵⁷).

Deterministic calculations have the advantage of being exact. However, if a model is complex, involving many possible events and intervening decisions based on those events, deterministic computations must exhaustively calculate the probability of every possible combination of events and decisions. In problems of even moderate complexity, this may involve millions of combinations. Stochastic models are, in essence, empirical samplings from these combinations, so that each combination appears in the final counts in proportion to its likelihood.

In addition to greater ease of complex simulation, stochastic models have the advantage of yielding not only average effects, but also measures of the uncertainty around the computed average (i.e., they can provide a confidence interval); deterministic calculations yield a point estimate only. A limitation of stochastic models is that complex simulations require intricate knowledge of the disease's natural history to estimate the parameters of the simulation. In the absence of such knowledge, the analyst must make the most reasonable assumptions about disease history. In either deterministic or stochastic modeling, lack of knowledge may be addressed through sensitivity analyses—varying the model parameters through reasonable ranges to observe the effect on the results. Generally it is desirable to use the simplest model possible, for which critical data are available for describing the parameters and their interrelationships.

Other Issues in Modeling to Estimate Effectiveness

Several issues germane to modeling will be briefly reviewed in this section, including specification of param-

ters and modeling patient characteristics, use of disease-specific or total mortality data, using models to “correct” for lead time and length biases, and model validation.

Specification of Survival Parameters and Modeling Patient Characteristics. Clinical trials and observational studies typically provide estimates of risk reduction or relative risk during the follow-up period, but give little indication how to estimate the survival curve for individuals beyond the end of the trial. Moreover, a trial restricted to a particular demographic or clinical group begs the question of what the effect might be in persons of younger or older ages, persons of the opposite sex, or persons with comorbidities. Thus, the analyst must make assumptions regarding the appropriate basis for extrapolation beyond the period of observation and to populations with different survival curves. For example, in a CEA comparing two thrombolytic therapies for acute myocardial infarction, the analysts used primary data from an RCT to estimate 1-year survival and then extended the observation period by modeling survival based on a separate database of patients with coronary heart disease.⁵⁵

The simplest assumption to make is that the age- and sex-specific risk of death for the affected population is modified by the disease in question, the intervention being evaluated, and any comorbidities that affect survival relative to the general population. A key choice is whether these three effects are additive or multiplicative.

Event probabilities in CEA models are often represented as conditional on patient characteristics, including age, gender, risk factors, stage of disease, and prior morbid events.^{58–62} These probabilities can be estimated separately for relevant subpopulations when data permit, but more often they are specified by an equation derived assuming a statistical relation between event probabilities and patient characteristics. The predictive equations can be derived using logistic regression, Poisson regression, proportional hazards models, or Bayesian analysis, to name a few techniques. These analyses can assume independence among the characteristics, or they may allow for interactions (e.g., effect modification); they can also be additive or multiplicative. Proportional hazard models and the logistic regression models are both essentially multiplicative. The declining exponential approximation to life expectancy (DEALE) model uses an additive function of risk factors.^{63,64} The implications of the choice of an additive versus a multiplicative assumption can be striking.^{9,10,47,65–68} When practical, sensitivity analysis should be used to evaluate different assumptions about parameter form.

Disease-Specific and All-Cause Mortality. Estimating length of life is a central problem in CEA. The main end point in many trials is disease-specific mortality—that is, mortality due to the disease addressed by the trial. However, the disease-specific mortality may be only part of the picture. For example, in estimating the effectiveness of cholesterol-lowering drugs to reduce deaths from cardio-

vascular disease, use of cardiovascular disease-specific mortality will overstate effectiveness if the intervention also leads to a higher rate of death from other causes.^{69,70}

Another caveat regarding the use of disease-specific mortality to estimate effectiveness in a CEA concerns misclassification.^{71,72} In RCTs, in which the investigators have drawn careful protocols for attribution of cause of death and can make this determination in follow-up of study participants, the attributed disease-specific mortality rates may be useful inputs to the CEA modeling process. However, in less-controlled studies, there may be either an underreporting or overreporting of disease-specific causes. Thus, it is suggested that CEAs use all-cause mortality as the basis for estimating life expectancy gains.

Use of Modeling to Address Lead Time and Length Biases.

Two related types of bias—lead time and length—make it difficult to determine with certainty if a screening intervention is effective in improving outcomes seen in nonrandomized controlled studies;^{15,73–75} modeling can be used to “correct” estimates of effectiveness for these biases. Lead time in a screening program is the time, in the normal course of disease, between the average time of early diagnosis by screening or case finding, and the average time of diagnosis in the absence of screening. Lead time bias is an overestimate of the increased survival associated with screening, owing to the fact that the disease is diagnosed earlier in its natural history. In extreme cases, all of the observed increase in survival with screening may be attributable to lead time bias, and there may be no actual prolongation of life.^{74,75} Lead time may have another consequence that is important for effectiveness estimation: earlier diagnosis and treatment afforded by screening may expose the patient to a longer period of adverse treatment effects than would occur in the absence of screening.

Length bias refers to the tendency for slower-growing, less-virulent disease to be detected in a screening program more often than more aggressive disease. As a result, those with aggressive disease are underrepresented in screened populations, and patients detected by screening may do better than unscreened patients, regardless of whether screening actually influences outcome, leading to an overestimation of effectiveness.^{15,74}

The analyst can address these biases by modeling the disease process directly; this requires estimates of such variables as tumor progression, stage-specific screening sensitivity, and stage-specific treatment response. One simulation model that incorporates such disease process modeling to avoid lead time and length biases is the MISCAN model.⁵⁷ For examples of this, and other approaches, the reader is referred to other sources.^{49,58–62,76–79}

Model Validation. Models often cannot be validated directly. The results of a model should, therefore, be accompanied by sensitivity analyses identifying which model inputs and parameters exert the most leverage on outputs. However, some aspects of models, such as which vari-

ables are included in the model inputs and their relations within the model's structure (i.e., additivity or multiplicity), are not easily amenable to sensitivity analysis. Thus, validation of a model may have to rest largely on the inherent reasonableness of the model and its assumptions. The technical accuracy of models must also be verified, to ensure that the model performs the calculations correctly as claimed. Computer programming and data entry errors, as well as logical inconsistencies in model specification, can all lead to errors that should be detected in the verification process. Reports based on models should contain assurances that the model has been verified.

The predictive validity of a model should also be evaluated when data are available to validate intermediate or final numerical predictions. For instance, predictions of cancer models can be compared with data on observed patterns in cancer incidence, staging, and mortality.⁸⁰ It is incumbent on the modeler to provide for the possibility of peer review and replication by colleagues; procedures for review of complex models need to be developed.

In conclusion, the goals of this article were to provide researchers and clinicians with a familiarity with the issues and terminology involved in evaluating the effectiveness of medical practices. The calculation of net effectiveness for CEA involves accounting for the many complex events that follow from the decision to intervene with patients or populations. Estimates of net effectiveness will rarely be based on the results from a single study that collects all of the outcomes of interest for the alternatives to be analyzed. Much more frequently, the process of estimating effectiveness will be one of constructing models that combine diverse information from several sources. Although we may see more future studies that are designed specifically to collect the primary data needed for a CEA,⁸¹ we are likely to continue to rely on approaches that combine data in mathematical models.

The members of the Panel on Cost-Effectiveness in Health and Medicine include Norman Daniels, PhD; Dennis G. Fryback, PhD; Alan M. Garber, MD, PhD; Marthe R. Gold, MD, MPH; David H. Hadorn, MD; Mark S. Kamlet, PhD; Joseph Lipscomb, PhD; Bryan R. Luce, PhD; Jeanne S. Mandelblatt, MD, MPH; Willard G. Manning, Jr., PhD; Donald L. Patrick, PhD; Louise B. Russell, PhD; Joanna E. Siegel, ScD; George W. Torrance, PhD; and Milton C. Weinstein, PhD.

We thank Kristine I. McCoy, MPH, for research, coordination, and editorial assistance in association with this project, and William F. Lawrence, MD, MSc, for editorial review of this manuscript.

REFERENCES

1. Torrance GW, Feeny D. Utilities and quality-adjusted life years. *Int J Technol Assess Health Care*. 1989;5:559-78.
2. Patrick DL, Erickson P. *Health Status and Health Policy*. New York, NY: Oxford University Press; 1993.
3. Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. *Cost-Effectiveness in Health and Medicine: Report of the Panel on Cost-Effectiveness in Health and Medicine*. New York, NY: Oxford University Press; 1996.
4. Gold MR, Patrick DL, Torrance GW, et al. Identifying and valuing outcomes. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. *Cost-Effectiveness in Health and Medicine: Report of the Panel on Cost-Effectiveness in Health and Medicine*. New York, NY: Oxford University Press; 1996.
5. Luce BR, Manning WG, Siegel JE, Lipscomb J. Estimating costs in cost-effectiveness analysis. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. *Cost-Effectiveness in Health and Medicine: Report of the Panel on Cost-Effectiveness in Health and Medicine*. New York, NY: Oxford University Press; 1996.
6. Park RE, Fink A, Brook RH, et al. Physician ratings of appropriate indicators for six medical and surgical procedures. *Am J Public Health*. 1986;76:766-72.
7. Leape LL. Practice guidelines and standards: an overview. *Qual Rev Bull*. 1990;16(2):42-9.
8. Leape LL, Brook RH. RAND Corporation Appropriateness Rating Program: Context and Purpose. Paper presented at Workshop to Improve Group Judgement for Medical Practice and Technology Assessment sponsored by the Council of Health Care Technology, Division of Health Care Services, Institute of Medicine, Washington, DC, May 15-16, 1990.
9. Kuntz KM, Weinstein MC. Life expectancy biases in clinical decision modeling. *Med Decis Making*. 1995;15:158-169.
10. Russell LR, Taylor WC, Jagannathan R, Milan E. What statistical model best describes heart disease risk? Evidence from the NHANES I epidemiologic follow-up study. In: *The Risk and Risk Factor Modeling Project: Final Report*. Washington, DC: Agency for Health Care Policy and Research, U.S. Public Health Service; 1995. Grant HS 07002.
11. Fryback DG. Bayes' theorem and conditional non-independence of data in medical diagnosis. *Comput Biomed Res*. 1978;11:423-34.
12. Poses RM, Cebul RD, Centor RM. Evaluating physicians' probabilistic judgments. *Med Decis Making*. 1988;8:233-40.
13. Ellenberg JH. Selection bias in observational and experimental studies. *Stat Med*. 1994;13(5-7):557-67.
14. Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics*. 1983;39:207-15.
15. U.S. Preventive Services Task Force (USPSTF). *Guide to Clinical Preventive Services*. 2d ed. Baltimore, Md: Williams and Wilkins; 1995.
16. Mandelblatt JS, Fryback DG, Weinstein MC, Russell LB, Gold MR, Hadorn DC. Assessing the effectiveness of health interventions. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. *Cost-Effectiveness in Health and Medicine: Report of the Panel on Cost-Effectiveness in Health and Medicine*. New York, NY: Oxford University Press; 1996.
17. Weiss NS. Control definition in case-control studies of the efficacy of screening and diagnostic testing. *Am J Epidemiol*. 1983;116:457-60.
18. Morrison AS. Case definition in case-control studies of the efficacy of screening. *Am J Epidemiol*. 1982;115:6-8.
19. Conner RJ, Prorok PC, Weed DL. The case-control design and the assessment of the efficacy of cancer screening. *J Clin Epidemiol*. 1991;44:1215-21.
20. Gill TM, Horwitz RI. Evaluating the efficacy of cancer screening: clinical distinctions and case-control studies. *J Clin Epidemiol*. 1995;48:281-92.
21. Selby JV, Friedman GD, Quesenberry CP Jr, Weiss NS. Case-control evaluation of screening. *J Clin Epidemiol*. 1996;49:390-1. Letter.
22. Selby JV. A case-control study of screening sigmoidoscopy and mortality from colorectal cancer. *N Engl J Med*. 1992;326:653-7.
23. Mosteller F. *Assessing Medical Technologies*. Institute of Medicine, Committee for Evaluating Medical Technologies in Clinical Use. Washington, DC: National Academy Press; 1985.
24. Mitchell JB, Bubolz T, Paul JE, et al. Using medicare claims for outcomes research. *Med Care*. 1994;32(suppl):JS38-51.

25. Massey JT, Moore TF, Parsons VL, Tadros W. Design and estimation for the National Health Interview Survey, 1985-94. *Vital Health Stat* 2. 1989;110.
26. Ezzati TM, Massey JT, Waksberg J, Chu A, Maurer KR. Sample design: Third National Health and Nutrition Examination Survey. *Vital Health Stat* 2. 1992;113:1-35.
27. Ingram DD, Makuc DM. Statistical issues in analyzing the NHANES I epidemiologic follow up study. *Vital Health Stat* 2. 1994;121:1-30.
28. Dickersin K, Berlin JA. Meta-analysis: state-of-the-science. *Epidemiol Rev*. 1992;14:154-76.
29. Hasselblad V, McCrory DC. Meta-analytic tools for medical decision making: a practical guide. *Med Decis Making*. 1995;15:81-96.
30. Sacks HS, Berrier J, Teitman D, Ancona-Berk VA, Chalmers TC. Meta-analysis of randomized controlled trials. *N Engl J Med*. 1987;316:450-5.
31. L'Abbe KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. *Ann Intern Med*. 1987;107:224-33.
32. Pettiti DB. Meta-analysis, Decision Analysis, and Cost-Effectiveness Analysis: Methods for Quantitative Synthesis in Medicine. New York, NY: Oxford University Press; 1994.
33. Eddy DM, Hasselblad V, Schachter R. Meta-analysis by the Confidence Profile Method: The Statistical Synthesis of Evidence. Boston, Mass: Academic Press; 1992.
34. Brophy JM, Joseph L. Placing trials in context using Bayesian analysis. GUSTO revisited by Reverend Bayes. *JAMA*. 1995;273:871-5.
35. Klevit HD, Bates AC, Castaneres T, Kirk EP, Sipes-Metzler PR, Wopat R. Prioritization of health care services. A progress report by the Oregon Health Services Commission. *Arch Intern Med*. 1991;151:912-6.
36. Banta HD, Luce BR. Health Care Technology and Its Assessment. Oxford, U.K.: Oxford Medical Publications; 1993.
37. Office of Technology Assessment (OTA), U.S. Congress. Identifying Health Technologies That Work: Searching for Evidence. OTA-H-608 Washington, DC: U.S. Government Printing Office; 1994; Five Background Papers. BP-H-142. Washington, DC: U.S. Government Printing Office; 1995.
38. Phelps CE, Mushlin AI. Focusing technology assessment using medical decision theory. *Med Decis Making*. 1988;8:279-89.
39. Raiffa H. Decision Analysis: Introductory Lectures on Choices Under Uncertainty. Reading, Mass: Addison-Wesley Publishing Co.; 1968.
40. Weinstein MC, Fineberg HV, Frazier AS, Neuhauser D, Neutra RR, McNeil BJ. Clinical Decision Analysis. Philadelphia, Pa: W.B. Saunders Co.; 1980.
41. Sox H, Blatt MA, Higgins MC, Marton KI. Medical Decision Making. Boston, Mass: Butterworths; 1988.
42. Lieu TA, Cochi SL, Black SB, et al. Cost-effectiveness of a routine varicella vaccination program for US children. *JAMA*. 1994;271:375-81.
43. Willems JS, Sanders CR, Riddiough MA, Bell JC. Cost-effectiveness of vaccination against pneumococcal pneumonia. *N Engl J Med*. 1980;303:553-9.
44. Riddiough MA, Sisk JE, Bell JC. Influenza vaccination: cost-effectiveness and public policy. *JAMA*. 1983;249:3189-95.
45. Mandelblatt JS, Wheat ME, Monane M, Moshief R, Hollenberg J, Tang J. Breast cancer screening for elderly women with and without co-morbid conditions: a decision model. *Ann Intern Med*. 1992;116:722-30.
46. Beck JR, Pauker SG. The Markov process in medical prognosis. *Med Decis Making*. 1983;3:419-58.
47. Weinstein MC, Coxson PG, Williams LW, Pass TM, Stason WB, Goldman L. Forecasting coronary heart disease incidence, mortality, and cost: the coronary heart disease policy model. *Am J Public Health*. 1987;77:1417-26.
48. Wong JB, Sonnenberg FA, Salem DN, Pauker SG. Myocardial revascularization for chronic stable angina. *Ann Intern Med*. 1990;113:852-71.
49. Eddy DM. Screening for breast cancer. *Ann Intern Med*. 1989;111:389-99.
50. Fahs M, Mandelblatt JS, Schechter C, Muller C. The costs and effectiveness of cervical cancer screening in the elderly. *Ann Intern Med*. 1992;117:520-7.
51. Krahm MD, Mahoney JE, Eckman MH, Trachtenberg J, Pauker SG, Detsky AS. Screening for prostate cancer: a decision analytic view. *JAMA*. 1994;272:773-80.
52. Tosteson AN, Rosenthal DI, Melton LJ, Weinstein MC. Cost effectiveness of screening perimenopausal white women for osteoporosis: bone densitometry and hormone replacement therapy. *Ann Intern Med*. 1990;113:594-603.
53. Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. *Med Decis Making*. 1993;13:322-38.
54. Paltiel AD, Kaplan EH. Modeling zidovudine therapy: a cost-effectiveness analysis. *J Acquir Immune Defic Syndr*. 1991;4:795-804.
55. Mark DB, Hlatky MA, Califf RM, et al. Cost effectiveness of thrombolytic therapy with tissue plasminogen activator as compared with streptokinase for acute myocardial infarction. *N Engl J Med*. 1995;332:1418-24.
56. Eddy DM. A computer-based model for designing cancer control strategies. In: Greenwald PG, Sondik EJ, eds. Cancer Control Objectives for the Nation: 1985-2000. Bethesda, Md: National Institutes of Health; 1986. NIH publication no. 86-2880, no. 2.
57. Habbema JDF, van Oortmarssen GJ, Lubbe JTHN, van der Maas PJ. The MISCAN simulation program for the evaluation of screening for disease. *Comput Methods Programs Biomed*. 1984;20:79-93.
58. Feuer EJ, Wun LM. How much of the recent rise in breast cancer incidence can be explained by increases in mammography utilization? A dynamic population model approach. *Am J Epidemiol*. 1992;136:1423-36.
59. Chang P. A simulation study of breast cancer epidemiology and detection since 1982: the case for limited malignant potential lesions. PhD dissertation. Department of Industrial Engineering, University of Wisconsin-Madison, 1993.
60. Eddy DM. Screening for Cancer: Theory, Analysis and Design. Englewood Cliffs, NJ: Prentice-Hall; 1980.
61. Eddy DM. Screening for cervical cancer. *Ann Intern Med*. 1990;113:214-6.
62. Eddy DM. Screening for colorectal cancer. *Ann Intern Med*. 1990;113:373-84.
63. Beck JR, Pauker SG, Gottlieb JE, Klein K, Kassirer JP. A convenient approximation of life expectancy (the "DEALE"), II: use in medical decision-making. *Am J Med*. 1982;73:889-97.
64. Keeler E, Bell R. New DEALEs: other approximations of life expectancy. *Med Decis Making*. 1992;12:307-11.
65. Weinstein MC, Stason WB. Hypertension: A Policy Perspective. Cambridge, Mass: Harvard University Press; 1976.
66. Goldman L, Weinstein MC, Goldman PA, Williams LW. Cost-effectiveness of HMG-CoA reductase inhibition for primary and secondary prevention of coronary heart disease. *JAMA*. 1991;265:1145-51.
67. Oster G, Epstein AM. Cost-effectiveness of antihyperlipemic therapy in the prevention of coronary heart disease. *JAMA*. 1987;258:2381-7.
68. Psaty BM, Koepsell TD, Manolio TA, et al. Risk ratios and risk differences in estimating the effect of risk factors for cardiovascular disease in the elderly. *J Clin Epidemiol*. 1990;43:961-70.
69. Muldoon MF, Manuck SB, Matthews KA. Lowering cholesterol concentrations and mortality: a quantitative review of primary prevention trials. *BMJ*. 1990;301:309-14.
70. Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Summary of the second report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel II). *JAMA*. 1993;269:3015-23.
71. Brown ML, Brauner C, Minnotte MC. Noncancer deaths in white adult cancer patients. *J Natl Cancer Inst*. 1993;85:979-87.
72. Greenfield S, Aronow HU, Elashoff RM, Watanabe D. Flaws in

- mortality data: the hazards of ignoring co-morbid disease. *JAMA*. 1988;260:2253-5.
73. Prorok PC, Hankey BF, Bundy BN. Concepts and problems in the evaluation of screening programs. *J Chron Dis*. 1981;34:159-71.
74. Morrison AS. *Screening in Chronic Disease*. 2d ed. New York, NY: Oxford University Press; 1992.
75. Black WC, Welch HG. Advances in diagnostic imaging and over-estimations of disease prevalence and the benefits of therapy. *N Engl J Med*. 1993;328:1237-43.
76. Shwartz M. A mathematical model used to analyze breast cancer screening strategies. *Operation Res*. 1978;26:937-55.
77. Shwartz M. Validation and use of a mathematical model to estimate the benefits of screening younger women for breast cancer. *Cancer Detect Prev*. 1981;4:595-601.
78. van Oortmarssen GJ, Habbema JD, van der Maas PJ, et al. A model for breast cancer screening. *Cancer*. 1990;66:1601-12.
79. Black WC, Welch HG. A Markov model of early diagnosis. *Acad Radiol*. 1996;3:S10-12.
80. Eddy DM. The frequency of cervical cancer screening: comparison of a mathematical model with empirical data. *Cancer*. 1987;60:1117-22.
81. Drummond MF, Davies L. Economic evaluation alongside clinical trials. *Int J Technol Assess Health Care*. 1991;7:561-73.