ECONOMICS AND HEALTH SYSTEMS RESEARCH
*SECTION EDITOR*
PETER G. DUNCAN

REVIEW ARTICLE

# Medical Technology Assessment: An Overview

Lee A. Fleisher, MD*, Srinivas Mantha, MD†‡, and Michael F. Roizen, MD‡

*Department of Anesthesiology and Critical Care Medicine, Johns Hopkins Medical Institutions, Baltimo
†Department of Anesthesiology and Intensive Care, Nizam's Institute of Medical Sciences, Hyderabad, In
‡Department of Anesthesiology and Critical Care Medicine, The University of Chicago, Chicago, Illinois

In attempting to address the issue of medical technology assessment, it is important to define technology. There are numerous definitions of technology. From the broadest perspective, John Kenneth Galbraith defined it as organized knowledge. The United States Congress Office of Technology Assessment defines medical technology as "the techniques, drugs, equipment, and procedures used by healthcare professionals in delivering medical care to individuals, and the systems within which such care is delivered." Assessment of medical technology therefore is the process of examining and reporting properties of a medical technology, such as safety, efficacy, feasibility, effectiveness, and indications for use. The past century has witnessed an explosion of new medical technologies. Once, medical care was relatively inexpensive because interventions were rare and personnel costs were low. As treatments became effective, new drugs and operations were integrated into routine medical practice, increasing the costs of medical care. Increased usefulness and cost led to the development of medical insurance to pay for new technologies. Payment of traditional medical insurance was retrospective; i.e., physicians and hospitals were reimbursed for procedures that had already been performed. Retrospective payment fueled the development of technology because increased use of technology led to increased reimbursement (Fig. 1), and technology, especially in the surgical arena, was left to develop unfettered by scientific proof. Every innovative procedure that could be conceived and tried was reimbursed in the 1960s and 1970s.

Recently, there have been dramatic changes in the method of medical reimbursement. Increasingly, medical bills are reimbursed on a prospective payment system, involving disease-related groups (Diagnosis-Related Groups [DRG]), accepted procedure codes (Current Procedural Terminology [CPT
tated payments (1–3). Increased use of t
innovations does not yield reimburse
that for existing DRG procedures and c
ogies that add cost without benefit re
for others in the medical systems. Co
nologies that decrease overall cost ar
Although, in theory, technology innov
expected to slow under the capitated s
research is required to determine wheth
be transformed into practice. Assessm
technology has emphasized the evalu
technologies for both outcomes and co
of making policy decisions for clin
reimbursement.

## Approval of Medical Technolo

Virtually all countries have some for
process before new drugs or devices ca
public, although their exact proces an
differ. In the United States, the Food a
ministration (FDA) is responsible for a
drugs and devices used for medical c
several excellent reviews on the FDA p
cess; therefore, the topic is discussed
herein (4,5).

The safety and efficacy of a new
demonstrated before its approval. The
insert) can be written so that it guide
administering or prescribing a drug w
able benefit to risk ratio. Classically, the
process has three phases. In the last dec
and III have been merged, and phase IV
ing) studies are now frequently required
approval. Phase I trials seek to find the
evaluate toxicity with pharmacodynam
macokinetics. These are usually trial
drugs in healthy volunteers. In pat
vanced disease or conditions that do
other therapies, toxic med ations, su
motherapeutic drugs, may be used in

Accepted for publication September 22, 1998.
Address correspondence and reprint requests to Lee A. Fleisher, MD, The Johns Hopkins Hospital, 600 N. Wolfe St., Carnegie 442, Baltimore, MD 21287. Address e-mail to lfleishe@welchlink.welch.jhu.edu.


0003-2999/98/$5.00

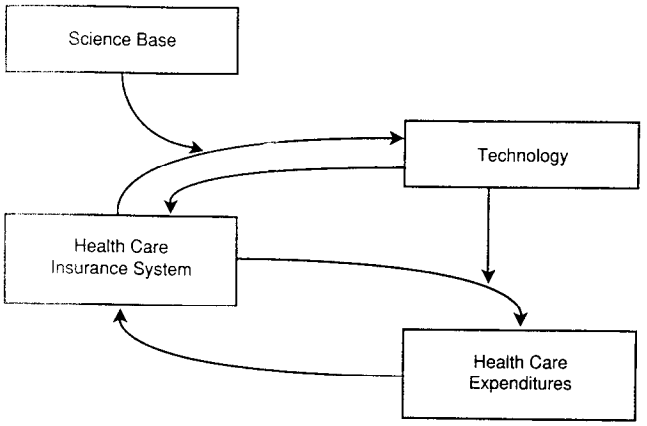


Figure 1. Driving force for technology development. Traditionally, a new technology is developed from a basic science discovery. This new technology would be reimbursed in a retrospective payment system, which further fuels development of other new technologies. In the prospective payment system (e.g., Diagnosis-Related Groups), new technologies do not lead to increased reimbursement, and only those technologies that are proven to be cost-effective will be adopted. Adapted from Weisbrod B. The nature of technological change: incentives matter! In: Adopting new medical technology, Washington, DC: National Academy Press, 1994.

Phase II studies are traditionally small (<300 patients), carefully controlled clinical trials in a tightly defined patient population designed to determine the effectiveness of a drug for a particular indication. Phase II studies focus on efficacy, common short-term side effects, and risks associated with a drug.

Phase III studies are expanded controlled and uncontrolled trials of effectiveness in less tightly controlled patient populations. The studies gather additional information about effectiveness and safety and involve from a few hundred to several thousand subjects; patients in Phase III studies often have comorbid conditions. A drug's effectiveness, pharmacokinetics, and pharmacodynamics are defined with such disease entities.

There are different requirements for FDA approval of drugs and devices. Traditionally, the primary requirement for approval of a new device was safety. Only recently has there been a requirement to prove that a device performs its stated function. There are two distinct methods of demonstrating that a new device is safe. The first is to demonstrate that the device is substantially similar to an already approved device. Many monitoring devices are in this category. If a device cannot be shown to be substantially similar to an approved one, then a formal evaluation process must be performed. A recent example of a device that required and obtained formal evaluation is the BIS monitor, for which new guidelines were developed for a "depth of anesthesia" monitor (6).

Interestingly, surgical procedures do not require an official approval process. However, reimbursement plays a critical role in the adoption of a new surgical procedure. If a procedure is deemed experimental, many medical insurance policies will not cover that episode of care, although there is a great deal of interest in the legislatures to require coverage of patients enrolled in clinical trials. Medicare plays an important role in this area because the third-party payers frequently follow Medicare's lead in reimbursing a specific procedure under an existing CPT code and for a specific DRG. New CPT codes can be obtained, but their inclusion under a specific DRG is a longer bureaucratic process that encourages fitting new procedures into existing CPTs. Because existing CPTs do not recognize innovation with financial rewards, innovation may be discouraged.

Traditionally, the Health Care Finance Administration (HCFA) allowed local and regional carriers, e.g., Blue Cross/Blue Shield plans, to make coverage decisions without requiring full evaluations. In 1993, the HCFA reduced its number of carriers and has taken a more active role in coverage decisions, more frequently limiting expensive technologies by time or provider. For example, heart transplants are reimbursed by the Medicare program only at specific sites and often require that specific data be submitted by the provider of the service. Despite these limitations, new technologies could be widely adopted, greatly adding to the cost of the Medicare program. An example of the importance of evaluating new surgical procedures is lung reduction surgery. Several centers throughout the country began performing this procedure in selected patients and reported excellent results, but there was no comparison group to determine the procedure's true efficacy (7). Federal officials calculated that widespread use of lung-reduction surgery would cause a great financial burden on the HCFA (Medicare) and Veterans Administration. This conclusion led federal officials to place a moratorium on reimbursement for this procedure. After objections from the medical community, a large-scale randomized trial was organized to test the efficacy of lung-reduction surgery. The National Institutes of Health and Agency for Health Care Policy and Research funded the scientific evaluation, and the HCFA funded the clinical care. Enrollment for this trial has begun at 18 medical centers. It is estimated that 2600 patients with emphysema will be enrolled, half of whom will be randomized to surgical treatment. Enrollment will continue until at least the summer of 1999. This trial may serve as a model for future evaluations of new surgical procedures.

## Diffusion of Technology

Once a drug or device is approved, the goal of the manufacturer is to diffuse the technology into clinical practice to generate profits. Similarly, specific clinical practices may be championed by clinicians to advance

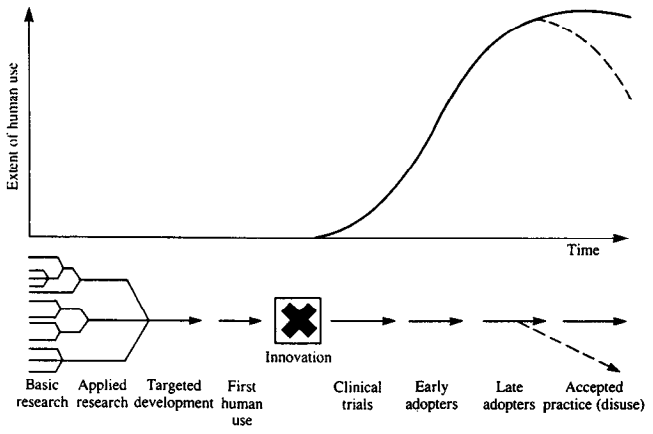


**Figure 2.** A model of the diffusion of a technology from basic science discoveries to clinical application. Once the discovery has a human use, it is considered an innovation. This first human use may involve a small study of its safety before trials to establish efficacy are begun. After its efficacy is established in clinical trials, there are both early and late adopters of the technology. After an newer technology is approved or the technology is demonstrated to be less effective, it may fall into disuse. Reproduced from the Office of Technology Assessment. Development of medical technology: opportunities for assessment. Washington, DC: US Government Printing Office, 1976.

an academic agenda. Frequently, the diffusion of the technology occurs faster than the formal evaluation.

Those who study diffusion of technology usually divide it into two groups: early and late adopters (Fig. 2). Once a drug or device is approved for sale, use slowly increases until it reaches a plateau, only to decrease when a better or less costly technology is introduced into clinical practice or the technology proves less effective in routine clinical practice.

Diffusion has changed dramatically in the modern communications era. Consumers are usually informed about the latest developments and may request a new drug soon after (or even before) FDA approval. Drugs used to treat human immunodeficiency virus may be the best example. There are numerous Internet Web sites devoted to the newest drugs and latest trials. Results of a new drug trial are frequently reported in the mass media before they are published in the medical literature. Therefore, the adoption of a drug may be quick (early adopters), but if the drug is less effective in routine clinical practice, its use may also decrease quickly. Even technologies associated with anesthesia care, such as pulmonary artery catheters and warming devices, have been described on television and in newspapers. Our patients may request that we not use (in the case of pulmonary catheters) or use (for warming devices) them.

Another recent change affecting diffusion is the direct marketing of prescription therapies, such as allergy medications, to consumers. This direct to the consumer advertising has become an issue in anesthesia care with the advertisement of outpatient surgery and anesthesia.

Government has played a dramatic role in the diffusion of one particular area of new technology: dialysis. Renal dialysis is an example of a technology that gets dramatic results: patients who otherwise would have died can be kept alive in a relatively good state of health. However, the costs for patients with renal failure who receive dialysis first and renal transplantation later are tremendous, and they are covered under the Medicare End-Stage Renal Disease plan. In 1995, Medicare's cost for this treatment of renal failure was estimated at well over $8 billion (8). No other disease group has ever been added to the Medicare plan, nor is it likely to be in the future.

In the Medical Resource Utilization Study (9) performed by the Rand Corporation, physicians reviewed charts for patients undergoing carotid endarterectomy, coronary angiography, and endoscopy. A panel of experts defined indications for each of the procedures and rated each procedure performed as appropriate, inappropriate, or equivocal; they concluded that as much as one third of the procedures performed were not indicated. Yet recent randomized clinical trials have found that some of the "unindicated procedures" according to the original Rand study would lead to improved outcome and would be indicated today (e.g., carotid endarterectomy for asymptomatic lesions >70%) (10). The concept of effectiveness is not static but may change as new studies are performed.

Another way to define appropriate utilization of a given surgical procedure is to study its use in different geographical locations. Wennberg et al. (11–13) found wide variability in the use of tonsillectomy in New England. They demonstrated wide variability in its use, which suggests that many of the procedures were performed for different indications in different locations.

Technologies may also be used if their perceived risk is low. For example, laparoscopic cholecystectomy is perceived to have less morbidity and lower costs than open cholecystectomy. When laparoscopic cholecystectomy was introduced, there was a dramatic increase in the number of cholecystectomies performed. The net result was that the expected reductions in costs of the procedure were offset by the increased number of procedures performed (14). Whether the additional procedures performed and advances in technology expanded the indications for surgery is not clear.

## Evidence-Based Medicine

Once the FDA approves a technology, or a procedure is approved for reimbursement, its use in clinical practice is frequently expanded beyond the approved indications, and it is often used in patients different from those studied in clinical trials. Additionally, in the approval process, the technology is rarely compared with a valid

alternative to determine whether it is cost-saving or whether any marginal increase in cost results in an improved outcome, i.e., its cost-effectiveness. Although some of this work is included in the approval process, most of these studies begin after approval is obtained, i.e., in the postmarketing phase.

Concern regarding the appropriate use of technology and financial constraints has led to the development of practice policies, guidelines, and utilization reviews by professional medical societies and by government organizations such as the Agency for Health Care Policy and Research. The guidelines define the strength of evidence for specific medical practices and identify the practices that should be abandoned. A full review of practice policies and their role in anesthesiology has recently been published (15).

As a result, outcomes-based research and evidence-based medicine have increased. Literature reviews and expert consensus panels form the basis of analysis; without a formal process, however, important references may be omitted, and occasionally studies of poor quality are accentuated, frequently reflecting the authors' bias on the subject. In contrast, evidence-based medicine refers to conscientious, explicit, and judicious use of the best evidence in making decisions about the care of individual patients (16). Recently, the Agency for Health Care Policy and Research devoted a multiday symposium on translating evidence into practice.

There are varying degrees of evidence based on the type of prospective study. Randomized, controlled clinical trials, which form the basis of FDA approval, are considered evidence of the highest quality because they define inclusion and exclusion criteria, treatment protocols, and outcomes of interest; they are usually either single or double-blinded (both patient and physician); and the randomization scheme and use of placebo (or accepted alternative treatments) ensure that the results are related to the intervention (internal validity). Importantly, these studies have a lower degree of external validity compared with less well controlled trials because the intervention or drug may not behave in the same manner when it is used in a more heterogenous population in whom treatment is not proscribed. Diffusion of the treatment into a heterogenous population distinguishes the drug efficacy under strict protocol versus effectiveness under real-world conditions. It has also led to large-scale clinical effectiveness trials in which care is governed much less by protocol. An example of this sort of trial is the current source study of remifentanil (16a). In this single-blinded (patient unaware), randomized, phase IV multicenter trial, 6000 patients received either a remifentanil- or fentanyl-based anesthetic to assess time to recovery, pain relief, and pharmacoeconomics. Although general guidelines for drug administration were given, the types of surgery and specific management decisions are left to the discretion of the caregivers. In this manner, the effectiveness of the drug in routine care can be determined.

A much weaker form of evidence is the prospective cohort study. In these studies, a group of patients is observed over time for specific outcomes. Such studies may be the only form of evidence available for a given intervention (17), and they are common in the evaluation of preoperative testing modalities. The quality of the evidence from prospective cohort studies depends on the accuracy of outcome assessment and the degree of blinding of the intervention. The American College of Physicians used quality ratings in the evaluation of preoperative testing studies and the development of guidelines (18).

Evidence gathered from nonprospective trials is considered still weaker. When a randomized, clinical trial cannot be performed on ethical or logistical grounds, a retrospective analysis is used. Clinical care may dictate inclusion of specific practices, even in the absence of good data to support their use. For example, a recent guideline stated that there is no evidence to support the use of pulmonary artery catheters in high-risk intensive care patients to reduce the risk of fluid and vasoactive therapies (19). A randomized clinical trial of right heart catheterization was aborted because clinicians considered it unethical to enroll critically ill patients (20). Case-controlled trials in which patients are matched for potential confounding risk factors to find a factor associated with a defined outcome offer weaker evidence than randomized trials. In one case-controlled study, the SUPPORT trial, patients with a pulmonary artery catheter were matched with patients without a catheter but who had the same aggregate propensity for receiving a catheter. Patients who received a pulmonary artery catheter were significantly more likely to die in the intensive care unit than those in the matched, equally ill cohort who did not (21). The major weakness of such trials is that they are highly dependent on the quality of the matching, for which the SUPPORT trial has been widely criticized (22). The results support the importance of and ethics for performing a randomized trial.

In some cases, primary outcomes are so rare or practice patterns are so diffuse that cohort trials would not contain a sufficient sample size. Administrative datasets or insurance claims data allow the researcher access to very large populations. Analysis of such data is retrospective from an evidence-based perspective. Examples of such datasets include the Medicare claims data (MEDPAR files), state discharge summaries, and insurance company claims data. Variations in practice and outcomes can be determined from these sources (18). Examples of retrospective analyses are those of the survival rates and utilization of resources

**Table 1.** Examples of Outcomes and Measures of Effectiveness

| |
|---|
| Mortality |
| Morbidity |
| Major |
| Myocardial infarction |
| Pneumonia |
| Pulmonary embolism |
| Minor |
| Nausea |
| Vomiting |
| Readmission |
| Patient satisfaction |
| Quality of life |

after an acute myocardial infarction (19,23). Retrospective analysis has also been used to determine the rates of morbidity after vascular surgery in an unselected population and their relationship to preoperative diagnostic testing and coronary interventions (24).

Although an analysis of administrative databases can assess low-frequency outcomes by examining large sample sizes and variations in clinical practices, there are significant limitations to the interpretation of the information. The data fields in an administrative database are usually few and dependent on the accuracy of coding, which limits the ability to assess comorbidities and nonfatal outcomes. There are frequently selection biases associated with the use of a particular technology, which cannot be accurately assessed. However, this type of analysis frequently provides sufficient evidence to generate hypotheses for future research.

## Outcomes

The potential benefits of any technology can be viewed as an improvement in morbidity and mortality, quality of life, or economics (Table 1). Medical outcomes have come under increasing scrutiny to measure so-called "hard outcomes" as opposed to surrogate end points, such as ST-segment change and postoperative nausea and vomiting (25,26). The value of any new technology must be judged against the importance of the changes in the outcome measure for both short- and long-term health. When evaluating any study or medical claim, one must consider whether the outcome measured is sufficiently important to change one's practice.

There is increasing interest in the value of technologic interventions for improving a patient's quality of life. Nonmedical outcome measures have been available in medicine for half a century. Karnofsky and Burchenal (27) proposed a functional assessment score for cancer that has been used in many therapy trials thereafter. Similarly, the New York Heart Association and the American Rheumatism Association have developed a functional score. From this initial approach of defining functional status developed the current approach to quality of life measures. There have been multiple definitions of health-related quality of life (HRQoL) proposed over the years. Schipper et al. (28) proposed that five basic tenets contribute to the current understanding of HRQoL: the psychological approach, the time trade-off utility concept, the community-centered concept, the reintegration concept, and the gap principle.

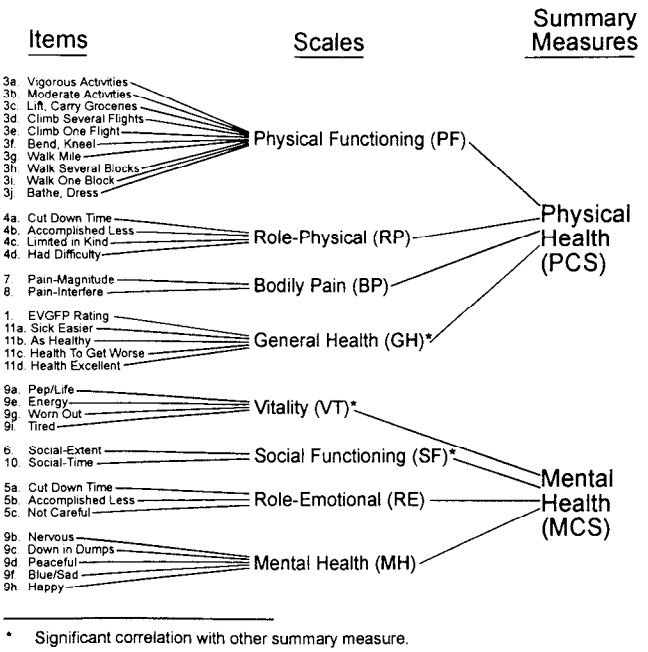


**Figure 3.** The model of the SF-36 quality of life model. EVGFP = excellent, very good, good, fair, poor; PCS = physical component summary; MCS = mental component summary. Reproduced with permission from Ware JE, Kosinski M, Keller SD. SF-36 physical and mental component summary measures: a user's manual. Boston: The Health Institute, 1994.

General measures of HRQoL include questions related to quality of life for multiple areas called domains. The SF-36 is a shorter version of the 245-item health status assessment questionnaire developed for use in the Medical Outcomes Study to evaluate the care of chronic medical and psychiatric conditions (29). The 36 items evaluate eight domains of physical and mental health, including functional ability, limitations in physical performance, bodily pain, anxiety and depression, sense of well-being, limitations to the fulfillment of emotional role requirements, social functioning, energy/fatigue levels, and perceived health status (Fig. 3). The SF-36 may be self-administered or used in interviews. It has been validated in numerous populations and shown to discriminate changes in health status (29). The scale also has 15 internal checks for consistency of respondent's answers. An even

smaller subset of items from the SF-36, the SF-12 encompasses the same eight dimensions as the SF-36 and has been used when resources are limited (30). For example, the SF-12 has been proposed as an instrument that could be administered by a nurse to all patients on admission and on discharge without overburdening the available patient care teams. Although these shorter instruments are easier to administer, they lose some of the original's discriminative ability.

Alternatively, specific scales have been developed for administration to patients with defined disease entities, such as those for rheumatoid arthritis and cardiovascular disease. These instruments are particularly useful in assessing the specific impact of an intervention.

HRQoL instruments have not been commonly used during the perioperative period because they were primarily developed to assess chronic healthcare states, as opposed to the acute changes seen perioperatively. However, the SF-36 or SF-12 is used to assess acute interventions, particularly with respect to cardiovascular and joint replacement surgery (31,32). They are becoming a *de facto* standard for assessing quality after many surgical procedures.

## Utilities and Quality-Adjusted Life Years

An alternative outcome used in modeling is utility. Utility is summarized by a single number along a scale ranging from 0 (death) to 1.0 (good health) (33). Years of life spent in a particular state multiplied by the respective utility value for that condition are called quality-adjusted life years (QALYs). The technique combines expected survival with expected quality to obtain a single number. For example, if the quality of life with a particular condition is estimated to be 0.5, spending 4 yr in that condition equals 2 QALYs (Fig. 4).

Although the perception of patients regarding health states is subjective, attempts have been made to quantify the value of quality of life. Essential techniques for quantification are linear scales, standard gamble, and time tradeoff methods. In the linear scale technique, a visual analog scale ranging from 0 (death) to 100 (perfect health) allows individuals to rate their health state. Although direct and simple, the method is not a time-utility measure; it does not force a choice under conditions of uncertainty and risk. The standard gamble method uses an economic utility theory and is based on the principle that a rational person will choose an option with the highest expected utility combining worth and probability of success. In this method, a subject is asked to imagine a hypothetical situation with two options: living in a diminished state of health (e.g., claudication) for life or taking a gamble (e.g., aortobifemoral surgery) with one of two outcomes: immediate restoration of normal health state

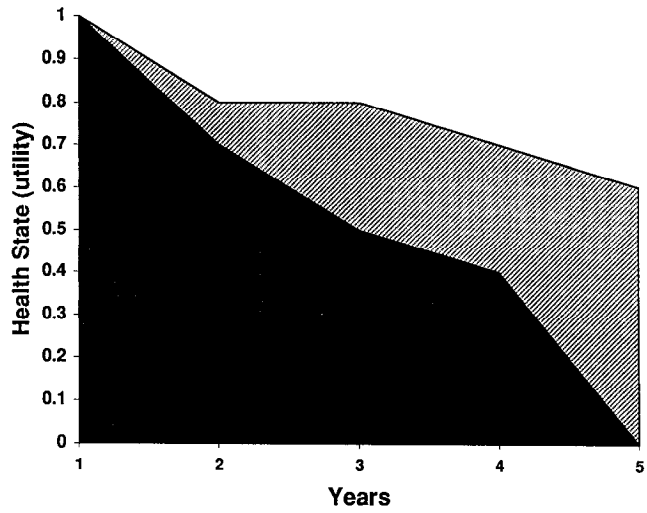


**Figure 4.** Improvements in quality-adjusted life years (QALYs) related to an intervention. The QALYs associated with a given intervention or disease are the sum of the health state (utility) for each given year of life, i.e., area under the curve. The effectiveness of a given intervention can be compared with the baseline state or alternative interventions by determining the differences in QALYs between the two states. ▨ = intervention, ■ = no intervention.

(with a probability of [*P*]) or death (with probability of [1 − *P*]). At first, a high probability of surgical success (99%) with a low surgical mortality (1%) is offered. The probabilities are varied correspondingly: 98% vs 2%, 95% vs 5%, 92% vs 8%, and so on until the "indifference" is reached. The point of indifference is the level of chance at which a patient would be indifferent to either claudication or surgery in this example. If a patient cannot decide between living with claudication and a 95% chance of perfect health after surgery versus a 5% chance of immediate death, then the utility of the particular health state is 0.95.

The time tradeoff method is a modification of the standard gamble technique but is easier to use. To obtain utility for a particular health state, the patient is asked to choose between life expectancy with a particular health state (severe claudication) versus life expectancy with normal health, assuming that in each case death would follow immediately. If the patient is willing to accept 9 yr of life with no claudication in exchange for 10 yr of life with claudication, then utility for the quality of life with claudication is 0.9 (9 divided by 10).

## Economics and Outcomes

In medical practice, one is often faced with a choice among competing models of managing disease states. Yet medical resources are limited and choices must be made between alternative uses of those resources. Savings from cost-effective strategies will allow the use of economic resources where they may best improve the quality of healthcare. The construction and reporting

of cost-effectiveness studies have been codified recently by a panel commissioned by the Public Health Service (34–36).

It is estimated that expenditures influenced (directly or indirectly) by anesthesia providers represent 3%–5% of the total healthcare costs of the United States (37). Economic analyses have been performed for anesthetic drugs and for additional perioperative issues, including routine preoperative laboratory testing, preoperative risk stratification strategies, monitoring techniques (e.g., pulmonary artery catheter), drugs to prevent or treat complications of anesthesia (postoperative nausea and vomiting), and chronic pain therapies. The term value-based anesthesia care, which includes technology and economic assessment, is used to describe the best patient outcome achievable at a reasonable economic input (38).

## *Types of Economic Analysis*

Cost-identification analysis determines the cost of an intervention (cost per service provided) when the outcomes of the interventions to be compared are equivalent (Table 2). The goal of the analysis is to identify the least expensive way of achieving the outcome. For example, if the outcome associated with the use of different neuromuscular blocking drugs in young healthy individuals is considered equivalent, then use of the less expensive pancuronium is a more economical strategy compared with the more expensive vecuronium. The limitation of this technique, also called cost-minimization analysis, is that, in concentrating on costs alone, differences in outcome and the economic effect on a patient may be neglected.

In cost-effectiveness analysis, the costs and results of alternative interventions are compared. This technique is used when the outcomes of different strategies are not equivalent but can be quantified. Ideally, quantification is expressed as a single measure, such as life years saved or infection averted. Different strategies can then be expressed in terms of cost per unit of outcome. The difference among incremental cost-effectiveness analysis, marginal cost-effectiveness analysis, and average cost-effectiveness ratio must be noted. The incremental cost-effectiveness ratio is an estimate of the economic cost per unit of effectiveness of switching from one intervention to another, e.g., selective noninvasive cardiac testing versus a test-all strategy. The issue is whether the additional improvement in effectiveness is worth the additional economic burden. The numerator of the ratio represents the difference in economic burden of an intervention and its alternative, whereas the denominator represents the difference in effectiveness (usually defined in terms of QALYs) of an intervention and its alternative.

**Table 2.** Dimensions of Economic Analysis

| Dimension | Issues for Consideration in the Analysis |
|---|---|
| The type of analysis | Cost identification or cost minimization |
| | Cost-effectiveness |
| | Cost-utility |
| | Cost-benefit |
| Types of costs and benefits | Direct medical |
| | Direct nonmedical |
| | Indirect morbidity and mortality |
| | Intangible |
| Perspective for analysis | Societal |
| | Patient |
| | Payer |
| | Provider |

Marginal cost-effectiveness analysis is useful when the scale of intervention increases, e.g., more visits per patient or expansion of ambulatory anesthesia services to accommodate a greater number of patients. When the net benefit per patient is likely to decrease, marginal cost-effectiveness analysis is needed. The term specifically refers to change in economic input and outcome of adding one unit of service.

The average cost-effectiveness ratio is estimated by dividing the cost of an intervention by a measure of effectiveness without regard to its competing alternatives, e.g., cost per infection detected for a particular screening strategy. The average cost-effectiveness ratio has limited value because it cannot be used to set priorities among competing strategies of medical technology.

Cost-utility analysis is a specific type of cost-effectiveness analysis in which the effects of the interventions are evaluated on both quantity and quality of life, thereby allowing health outcomes to be measured in terms of QALYs. QALYs combine a particular utility-based measure of quality of life with a quantitative measure of life years to obtain a single measure of lifetime utility. One QALY is equal to 1 yr at full health for that one individual.

In cost-benefit analysis, monetary value is placed on benefits or health outcomes. Both costs of providing care and effectiveness are measured in the same monetary unit, such as dollars. Results of a cost-benefit analysis could be expressed as the difference between economic input and benefit (expressed in monetary value, subtracting cost from benefit) or as the ratio of the two (ratio of benefit to cost). In general, calculating net benefit is preferred to calculation of the benefit to cost ratio.

Two methods are available for assigning monetary value to health outcomes: human-capital and willingness-to-pay methods. In the human-capital method, health outcomes are valued by an amount equivalent to the individual's contribution to the economy; i.e., the future earnings of the patient. With the human-capital approach, health benefits

for low-income patients, minorities, women, and the elderly may be valued less than benefits for others. Critics of the human-capital method raise the ethical question, "Does the value of human life depend on the earning potential of an individual?" As the name implies, the willingness-to-pay method estimates how much a society is willing to pay to provide an additional QALY. Although difficult to implement, many economists favor the willingness-to-pay method. In general, cost-benefit analysis is used less frequently than cost-effectiveness analysis or cost-utility analysis because of the difficulty in assigning economic value to health outcomes.

The terms used as surrogates for "cost" in the medical literature require definition. Charge is the amount that the hospital, clinic, physician, or pharmacy attempts to recover (or bills) for providing a service. Payment is the amount actually paid for the service by the individual or the third-party payer. Cost of a particular service is a function of all the resources consumed for that service. The types of relevant costs include direct, indirect, and intangible (39). Many analyses also calculate costs by multiplying charges by a fraction known as the cost-charge ratio reported to Medicare. Unfortunately, this type of calculation assumes that all costs are fixed.

Direct costs of medical care are the organizing and operational expenditures for its delivery. These costs can be medical or nonmedical. Direct medical costs include those incurred by hospitalization, drugs, and physicians and other relevant personnel. The time and motion method is often used to determine hospital costs, typically the direct costs of drugs. The time taken for a nurse to gather materials and to prepare and administer a medication is one example. Direct nonmedical costs include family and patient expenses that result from illness: food, transportation, family lodging, and home help. Direct nonmedical costs can be substantial and are not usually covered by insurance companies. Although these costs are not usually included in analyses, direct nonmedical costs should be included when a cost analysis is performed from a patient's perspective.

Indirect costs are the cost of loss of income (lost productivity) due to illness or death, including absence from work, lost wages, decreased earnings, and the need to change jobs. Indirect cost or productivity losses caused by an intervention should be contrasted with the indirect costs of illness. The indirect costs of an illness are usually measured by an extension of the human-capital approach. Reduction in the indirect costs of illness is often estimated as a monetary benefit, especially for cost-benefit studies.

Intangible costs represent the nonmonetary costs of illness, such as pain, suffering, and grief, expressed in monetary terms. These costs form a part of the denominator in cost-benefit analysis that uses the willingness-to-pay method. In cost-utility analysis, such items are not given a dollar value but are included in the determination of health outcomes, i.e., in the calculation of QALY. Therefore, even in the cost-utility analysis, intangible costs become a part of denominator.

An area of intense debate is the inclusion of potential payments to society. For example, should potential savings to Social Security be included in the cost analysis if someone dies from the treatment? This savings is rarely included in analyses.

An important aspect of cost determination is cost-finding, which describes the highly complex procedure for cost delineation. In cost delineation, costs are classified as fixed or variable. Fixed costs are the ongoing costs of providing service that are unrelated to volume. Salaries of operating room managers and costs of operating room and recovery room monitors are fixed costs. Variable costs vary as a function of the volume of service. Examples of variable costs in the operating room and recovery room include supplies, drugs, and supplemental nursing services.

### *Economic Models in Trials*

An economic analysis can be performed alongside a clinical trial or be included as part of a model. Data on costs for each intervention and outcome can be collected prospectively. A cost-effectiveness, cost-benefit, or cost-finding analysis can then be calculated directly between two interventions. Although a randomized, clinical trial or meta-analysis may provide some information about the efficacy of an intervention; the lack of cost data in most trials; the use of a placebo or control rather than a comparable, clinically relevant, alternative intervention; and outcome measures of mortality, myocardial infarction, and stroke as opposed to QALYs do not permit evaluation of cost-effectiveness. Additionally, technologies are quickly superseded by newer advances, sometimes before the evaluation of a technology is completed. Interventions such as screening and primary prevention are more difficult to evaluate by using a randomized trial design because of the long lag time between intervention and outcomes. Therefore, most cost-effectiveness analyses use a decision analysis model that combines data from multiple sources: individual or meta-analysis of clinical trials; natural history studies; hospital, regional, or national databases; and, when data are lacking, expert opinion (Delphi survey).

Decision analysis is an explicit analytic tool designed to facilitate complex clinical therapeutic or diagnostic decisions in which many variables must be considered simultaneously (40–42). It is important to distinguish between clinical decision analysis and policy applications of decision analysis. Clinical decision analysis is applied to individual patients and can be performed at the bedside. In policy applications, decision analysis is applied to society, populations, or

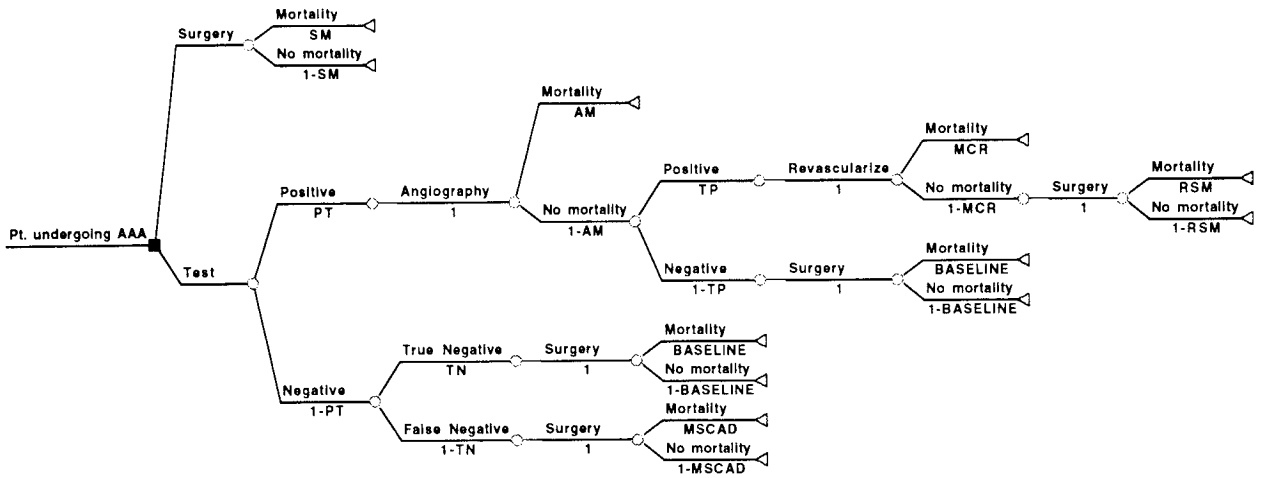


**Figure 5.** A representative decision algorithm evaluating the decision between vascular surgery alone or coronary artery revascularization before vascular surgery. There are currently no randomized trials to address the optimal strategy. By outlining the multiple decision points at which a patient can sustain mortality by choosing to undergo coronary revascularization first, the optimal strategy for preoperative evaluation can be demonstrated. Specifically, variation in mortalities at each decision point can change the optimal strategy. Pt = patient, AAA = abdominal aortic aneurysm, SM = surgical mortality, PT = positive test, AM = angiographic mortality, TN = true negative, TP = true positive, MSCAD = mortality for vascular surgery in patients with coronary artery disease, MCR = mortality from coronary revascularization, RSM = revascularized surgical mortality. Reproduced with permission from Fleisher LA, Skolnick ED, Holroyd KJ, Lehmann HP. Coronary artery revascularization before abdominal aortic aneurysm surgery: a decision analytic approach. Anesth Analg 1994;79:661–9.

groups of patients (43). The first step in decision analysis is constructing a decision tree (Fig. 5). A decision tree is a map of all relevant courses of action and their associated outcomes. The tree, built from left to right, consists of nodes, branches, and outcomes. A decision node is a branch point representing a diagnostic or therapeutic decision, conventionally depicted as a square. A branch point of a chance outcome not directly controlled by the physician is represented by a circle. Outcomes are depicted as rectangles or triangles. Events depicted at chance nodes are associated with probabilities. The probabilities of all events at a node must be equal to 1; i.e., chance nodes define alternative events that do not overlap. The outcome of each branch of the tree is a utility. Utility measures can be of various types; for example, 30-day postoperative survival (utility of 1 for survival and 0 for death), QALYs, or costs. Examples of utilities include mild angina of 0.89, severe angina with congestive heart failure of 0.78, and stroke of 0.4–0.6.

The expected utility of each potential course of action is a function of both the probability of the outcome and its utility. Multiplying the utilities by the probability of the outcome determines the expected utility of taking a particular action for each limb of a decision tree. The sum of the expected utilities of all the limbs gives the overall expected utility of a specified decision option. This process is called averaging. Choosing the optimal decision depends on the type of utility. The optimal decision is the one with the greatest overall expected utility for survival or QALYs and the lowest cost.

The ability to perform sensitivity and threshold analyses is one advantage of decision analysis. Sensitivity analysis assesses the impact of variations in probabilities and utilities on the final decision and the stability of the assumptions made in structuring the tree. Sensitivity analysis is performed by varying the assigned probabilities of one variable (one-way sensitivity analysis), two variables (two-way sensitivity analysis), or three variables (three-way sensitivity analysis) at a time. Threshold analysis defines specific assumptions at which a decision should be switched. Threshold values are revealed in one-way sensitivity analysis by the point at which the strategy with the greatest overall utility or lowest cost changes. Decision analysis has been applied to the question of whether preoperative testing and coronary revascularization should be performed before major vascular surgery (44). By using two-way sensitivity analysis to vary the probability of mortality from undergoing vascular surgery without testing and the mortality from coronary revascularization before vascular surgery (Fig. 6), the optimal strategy (least overall mortality) would vary among different institutions.

## Technology Assessment and Clinical Practice

Because the traditional literature review or consensus conference is not sufficient to provide an evidence-based perspective of practice, how does one synthesize the available literature? Meta-analysis has been defined as the quantitative summary of research in a particular area and the practice of using statistical methods to combine the outcome of a series of different experiments or investigations (45). As part of the selection criteria for the

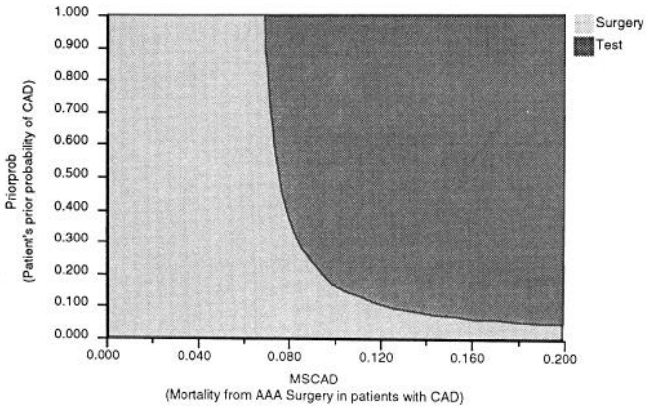


**Figure 6.** An example of a two-way sensitivity analysis based on the decision analysis proposed in Figure 5 demonstrating the optimal preoperative strategy of surgery alone or coronary revascularization before vascular surgery. Two of the critical variables in the decision analysis are varied within the clinically relevant range. As the probability of mortality from coronary revascularization increases, vascular surgery alone becomes the preferred strategy. In contrast, as the probability of mortality from aortic surgery in patients with significant coronary artery disease increases, coronary revascularization before vascular surgery becomes the optimal strategy. The average mortality for vascular surgery in patients with significant coronary artery disease is 9.5%, which suggests that the strategy with the lowest mortality is very sensitive to local morbidity and mortality. However, if long-term mortality is included in the model, coronary revascularization may be more beneficial. CAD = coronary artery disease, MSCAD = mortality for vascular surgery in patients with coronary artery disease, AAA = abdominal aortic aneurysm. Reproduced with permission from Fleisher LA, Skolnick ED, Holroyd KJ, Lehmann HP. Coronary artery revascularization before abdominal aortic aneurysm surgery: a decision analytic approach. Anesth Analg 1994;79:661–9.

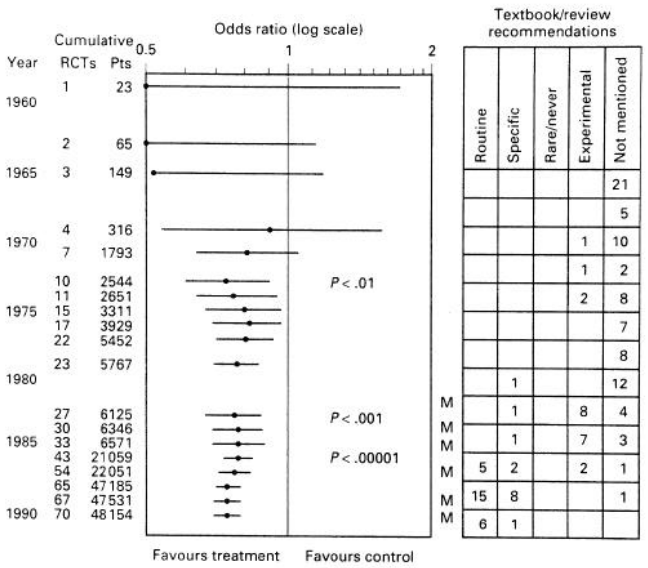


**Figure 7.** The effect of additional randomized clinical trials (RCTs) on a cumulative meta-analysis of the beneficial effects of thrombolytic therapy after an acute myocardial infarction. If the point estimate and confidence interval lies entirely to the left of 1.0, then there is a significant benefit to the intervention. The table to the right lists the number of recommendations to use thrombolytic therapy in textbooks. There is a long delay between the routine recommendation of its use and the point at which a meta-analysis would have demonstrated a statistically significant benefit. Pts = patients, M = meta-analysis published. Reproduced with permission from Antman EM, Lau J, Kupelnick B, et al. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts: treatments for myocardial infarction. JAMA 1992;268:240–8.

literature, the investigator should define the MEDLINE search words, the years included, and the secondary search of references from articles identified. There is debate in the literature regarding the appropriateness of including nonrandomized clinical trials. Some meta-analyses rate the quality of the literature and include only studies of fair to strong quality. The outcomes of interest may be recalculated to maintain consistency. For example, meta-analysis of perioperative cardiovascular studies may recalculate outcomes to include only myocardial infarction or death (46,47).

Through meta-analysis, sufficient evidence can be found to support or abandon a new technology or treatment before completion of large trials testing efficacy. One example is the use of thrombolytics for patients who sustain an acute myocardial infarction (48). In this case, the recommendations of the clinical experts lagged behind those which would have been advocated by a meta-analysis of the available literature (49) (Fig. 7).

Despite the benefits, the results of meta-analysis should be viewed with caution. The trials included in a meta-analysis are frequently based on those found in MEDLINE, which indexes >3900 biomedical journals published in the United States and 70 foreign countries. This represents only approximately one quarter of the world's literature, as non–English-language journals are not well represented. Retrieval is highly dependent on indexing. There is also significant bias toward publication of positive rather than negative trials. Finally, the data from the trials included in a meta-analysis are often from heterogeneous studies. Despite the statistical techniques used to minimize its effects, heterogeneity can still influence outcome.

## Codifying the Results of Technology Assessment

Practice policies or guidelines are the summation by clinicians of the available evidence about the benefits and risks of a treatment plan. Guidelines are a method of codifying recommendations regarding the use of a given technology (50). There are several types of recommendations that fall into the general category of a practice parameter. A standard implies that a therapy or practice should be used for patients with a particular condition. Standards are only approved if an assessment of the probabilities and utilities of the group indicate that the decision to choose the treatment or a strategy would be virtually unanimous. If a

particular therapy or strategy is considered a standard, it is cost-effective for those for whom it is recommended. Standards are intended to be applied rigidly. Guidelines are intended to be more flexible than standards, but they should be followed in most cases. Depending on the patient, setting, and other factors, guidelines can and should be tailored to fit individual needs. Like standards, guidelines should be cost-effective. Options are neutral with respect to use of an intervention or a strategy. Options merely note that different interventions are available and that different people make different choices. Options leave practitioners free to choose any course.

There is increasing interest in defining the value of both established and new medical technologies from an evidence-based approach. Many of the techniques and tools that have been developed in other specialties have been applied to the delivery of anesthetic care. It is important to understand the strengths and weaknesses of the methods used to evaluate medical technologies to determine when such technologies should be adopted into one's local practice and how to define their use to the increasing pressure from the third-party payers.

## References

1. Gay EG, Kronenfeld JJ. Regulation, retrenchment: the DRG experience—problems from changing reimbursement practice. Soc Sci Med 1990;31:1103–18.
2. Mitchell JB. Physician DRGs. N Engl J Med 1985;313:670–5.
3. Scarpaci JL. DRG calculation and utilization patterns: a review of method and policy. Soc Sci Med 1988;26:111–7.
4. Sanborn MD, Godwin HN, Pessetto JD. FDA drug classification system. Am J Hosp Pharm 1991;48:2659–62.
5. Myers A, Moore SR. The drug approval process and the information it provides. Drug Intell Clin Pharm 1987;21:821–6.
6. Gan TJ, Glass PS, Windsor A, et al. Bispectral index monitoring allows faster emergence and improved recovery from propofol, alfentanil.
7. Cooper JD, Patterson GA, Sundaresan RS, et al. Results of 150 consecutive bilateral lung volume reduction procedures in patients with severe emphysema. J Thorac Cardiovasc Surg 1996; 112:1319–29.
8. Garella S. The costs of dialysis in the USA. Nephrol Dial Transplant 1997;12:10–21.
9. Chassin MR, Kosecoff J, Park RE, et al. Does inappropriate use explain geographic variations in the use of health care services? A study of three procedures. JAMA 1987;258:2533–7.
10. Toole JF. ACAS recommendations for carotid endarterectomy [letter]. Lancet 1996;347:121.
11. Wennberg JE, Gittelsohn A. Health care delivery in Maine. I. Patterns of use of common surgical procedures. J Maine Med Assoc 1975;66:123–30,149.
12. Wennberg JE, Gittelsohn A, Soule D. Health care delivery in Maine. II. Conditions explaining hospital admission. J Maine Med Assoc 1975;66:255–61,269.
13. Wennberg JE, Gittelsohn A, Shapiro N. Health care delivery in Maine. III. Evaluating the level of hospital performance. J Maine Med Assoc 1975;66:298–306.
14. Legorreta AP, Silber JH, Costantino GN, et al. Increased cholecystectomy rate after the introduction of laparoscopic cholecystectomy. JAMA 1993;270:1429–32.
15. Shomaker TS. Practice policies in anesthesia: a foretaste of practice in the 21st century. Anesth Analg 1995;80:388–403.
16. Sackett DL, Rosenberg WM. On the need for evidence-based medicine. J Public Health Med 1995;17:330–4.
16a. Fleisher LA, Glass PSA, Roizen MA, Twersky R, Tuman K, Warner DS, Colopy M, Jamerson BD: Remifentanil provides superior hemodynamics and faster time to extubation than a fentanyl based anesthetic in a large scale trial of effectiveness. Anesthesiology 1998;89:A33.
17. Hornberger J, Wrone E. When to base clinical policies on observational versus randomized trial data. Ann Intern Med 1997; 127:697–703.
18. Palda VA, Detsky AS: Perioperative assessment and management of risk from coronary artery disease. Ann Intern Med 1997;127:313–28.
19. Anonymous: Practice guidelines for pulmonary artery catheterization. A report by the American Society of Anesthesiologists Task Force on Pulmonary Artery Catheterization. Anesthesiology 1993;78:380–94.
20. Guyatt G. A randomized control trial of right-heart catheterization in critically ill patients. J Intensive Care Med 1991;6:91–5.
21. Connors AF Jr, Speroff T, Dawson NV, et al. The effectiveness of right heart catheterization in the initial care of critically ill patients. JAMA 1996;276:889–97.
22. Sprung CL, Eidelman LA. The issue of a U.S. Food and Drug Administration moratorium on the use of the pulmonary artery catheter. New Horiz 1997;5:277–80.
23. Tu JV, Pashos CL, Naylor CD, et al. Use of cardiac procedures and outcomes in elderly patients with myocardial infarction in the United States and Canada. N Engl J Med 1997;336:1500–5.
24. Fleisher LA, Eagle KA, Shaffer T, Anderson G. Mortality after major vascular surgery: Analysis of the Medicare database [abstract]. Anesth Analg 1997;84:SCA43.
25. Fleisher LA. . . . But, is suppression of postoperative ST segment depression an important outcome? Anesth Analg 1997;84: 709–11.
26. Fisher DM. Surrogate end points: are they meaningful? Anesthesiology 1994;81:795–6.
27. Karnofsky DA, Burchenal JH. The clinical evaluation of chemotherapeutic agents in cancer. In: Macleod CM, ed. Evaluation of chemotherapeutic agents. New York: Columbia University Press, 1949.
28. Schipper H, Clinch JJ, Olweny CLM. Quality of life studies: definitions and conceptual issues. In: Spilker B, ed. Quality of life and pharmacoeconomics in clinical trials. Philadelphia: Lippincott-Raven, 1996:11–23.
29. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473–83.
30. Ware JE Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996;34:220–33.
31. Mangione CM, Goldman L, Orav EJ, et al. Health-related quality of life after elective surgery: measurement of longitudinal changes. J Gen Intern Med 1997;12:686–97.
32. Harlan WR III, Sandler SA, Lee KL, et al. Importance of baseline functional and socioeconomic factors for participation in cardiac rehabilitation. Am J Cardiol 1995;76:36–9.
33. Pauker SG, Pauker SP, Wong JB. The utility of thresholds and the threshold utility. Med Decis Making 1987;7:201–2.
34. Russell LB, Gold MR, Siegel JE, et al. The role of cost-effectiveness analysis in health and medicine: Panel on Cost-Effectiveness in Health and Medicine. JAMA 1996;276:1172–7.
35. Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses: Panel on Cost-Effectiveness in Health and Medicine. JAMA 1996;276:1339–41.
36. Weinstein MC, Siegel JE, Gold MR, et al. Recommendations of the Panel on Cost-Effectiveness in Health and Medicine. JAMA 1996;276:1253–8.
37. Johnstone RE, Martinec CL. Costs of anesthesia. Anesth Analg 1993;76:840–8.
38. Orkin FK. Moving toward value-based anesthesia care. J Clin Anesth 1993;5:91–8.

Clinical economics: a guide to the economic anal-
practices. JAMA 1989;262:2879–86.
ssirer JP. Decision analysis. N Engl J Med 1987;

oskowitz AJ, Lau J, Pauker SG. Decision analysis:
ort. Ann Intern Med 1987;106:275–91.
S, Detsky AS. Users' guides to the medical liter-
w to use a clinical decision analysis B. What are
l will they help me in caring for ny patients?
d Medicine Working Group. JAMA 1995;273:

er SG. Decision analysis for public health: prin-
trations. Annu Rev Public Health 1984;5:135–61.
olnick ED, Holroyd KJ, Lehmann HP. Coronary
ularization before abdominal aortic aneurysm
cision analytic approach. Anesth Analg 1994;79:

-analysis: current issues in r search synthesis. Stat
253–62.

46. Mantha S, Roizen MF, Barnard J, et al. Relative effectiveness of four preoperative tests for predicting adverse cardiac outcomes after vascular surgery: a meta-analysis. Anesth Analg 1994;79: 422–33.
47. Shaw LJ, Eagle KA, Gersh BJ, Miller DD. Meta-analysis of intravenous dipyridamole-thallium-201 imaging (1985 to 1994) and dobutamine echocardiography (1991 to 1994) for risk stratification before vascular surgery. J Am Coll Cardiol 1996;27: 787–98.
48. Stampfer MJ, Goldhaber SZ, Yusuf S, et al. Effect of intravenous streptokinase on acute myocardial infarction: pooled results from randomized trials. N Engl J Med 1982;307:1180–2.
49. Antman EM, Lau J, Kupelnick B, et al. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts: treatments for myocardial infarction. JAMA 1992;268:240–8.
50. Arens JF. A practice parameters overview. Anesthesiology 1993; 78:229–30.