Economic Endpoints in Clinical Trials

Mark A. Hlatky

New and effective medical therapies are constantly being developed and have led to steady improvements in the health and well-being of patients. However, these developments are one of the primary reasons that the cost of health care continues to increase. Rising health care costs have led policymakers to question whether all new treatments are worthwhile, especially expensive ones. As a result, data about the economic consequences of new therapies are now commonly collected in clinical trials so that cost-effectiveness can be assessed at the same time as clinical efficacy (1, 2). This new trend in clinical trials has stimulated methodological work on the data collection and analysis needed to perform an economic evaluation within a randomized trial (3–5). In this article, I first review methods for collecting economic data in clinical trials and then discuss particular issues that arise in data analysis. Finally, I discuss performing cost-effectiveness analyses based on data from clinical trials.

DATA COLLECTION

It is useful to think about the cost of a medical product or service as a function of both the resources used to produce it and the price of each of those resources. Thus, the total cost might be written as

\[ C_j = \sum_i P_i \times Q_{ij}, \]

where \( C_j \) = the cost for patient \( j \), \( P_i \) = the price of resource \( i \) used for patient \( j \), and \( Q_{ij} \) = the quantity of resource \( i \) used for patient \( j \).

If, for example, we wanted to determine the cost of a coronary angiogram for patient \( j \), we might identify the supplies used in the procedure; the amount of time spent by the physician, nurses, and technicians in performing the procedure; and the cost of using the catheterization laboratory facility with its expensive capital equipment. We need to set reasonable boundaries on this exercise to make it tractable, for the cost of each of these resources could easily be broken down in the same fashion (e.g., the cost of making the supplies used in the procedure could be calculated in turn by enumerating the cost of materials, labor, and capital equipment). In most clinical studies, cost finding is limited to those goods and services provided directly to the patient.

Two major methods have been used to measure costs in medical studies. The first starts with billing and financial records and establishes costs based on those data. The second starts with data about consumption of “key resources” and uses that information to establish cost. Each method has strengths and weaknesses.

Billing data

Most of the early clinical trials collected economic data by using the first method, taking advantage of very detailed billing records generated routinely by the fee-for-service medical system in the United States. Prior to the mid-1980s, physicians, hospitals, and other health care providers would care for an individual patient and then submit bills to health insurance companies, which in turn would pay the billed charges. Hospital bills were typically very detailed, listing the quantity of every test, medication, service, and procedure performed along with the charge (e.g., five chest radiographs at $50 each). Physicians submitted detailed bills for their services, as did pharmacies, laboratories, and other health care suppliers. In principle, the cost of care could be established simply by obtaining copies of all relevant billing records and summing the charges.

There are practical and theoretical limitations to using billing data in clinical trials. At the practical level, it is very difficult to obtain copies of a bill for every service, so data collection could be labor intensive at best, incomplete at worst. At a theoretical level, the charge for a service was often a poor reflection of its actual cost, since medical care is not provided in a truly competitive market. Thus, the charges for some medical services (e.g., coronary bypass surgery) were often marked up far in excess of their true costs to subsidize other services (e.g., the emergency department). Therefore, the charge for a service was not closely related to its cost (6).

The practical difficulty of obtaining a copy of every hospital and physician bill has been addressed by limiting data collection to those services that are either expensive or likely to be consumed differently by patients in the treatment groups within a clinical trial. Hospital admissions are much more expensive than any other medical service, so it is essential that hospitalization data collection be complete.
Physician fees for surgery and expensive laboratory tests or drugs are also examples of big-ticket costs that should be a priority for data collection.

The issue of distorted prices has been addressed by using standardized methods to convert charges to costs. Every year, hospitals in the United States must submit detailed cost reports to the Health Care Financing Administration. These public reports break down a hospital’s costs in specified categories and also list the charges billed in these same categories. There is a standard method for allocating overhead costs (e.g., maintenance, utilities) to the revenue-generating departments. The billed charges can be compared with the costs at the level of individual hospital departments. For example, if the radiology department cost $5 million to operate but incurred charges of $10 million, the cost-to-charge ratio would be \( \frac{5}{10} = 0.5 \). The cost-to-charge method converts charges to costs by using the appropriate ratios and summing over all departments. This approach has been used in many US trials, especially those performed in a small number of hospitals that can provide complete data about charges and costs.

**Resource consumption**

The second major approach to collecting economic data is to start with direct measures of resource consumption rather than bills. This method is easier to use in large studies and in settings in which patients are not billed for services. The research team must establish a priori the level of detail needed to compare the costs of therapy fairly. This list typically includes key resources such as hospitalization, invasive procedures, expensive medications, and physician services (especially for surgery). Data about use of these resources can then be collected on either the standard clinical data forms or separate data forms specifically designed for the economic portion of the study.

The quantity of key resources used by a patient provides the starting point for measuring costs. The next step is to establish the cost associated with each unit of resource consumption. One common technique is to use a standard price or fee schedule as set by payors. In the United States, Medicare reimbursements to hospitals and physicians are de facto national standard price schedules, and they represent the actual cost of care for a large segment of the population. Other countries (e.g., Australia) have established similar sets of prices for medical services, and these administrative price schedules have considerable face validity as appropriate cost weights for a randomized trial.

A less common method of establishing the cost of a medical service is to measure the actual cost of producing a service by using a time-motion study or other “microcosting” approach. This method is feasible in hospitals that have detailed computerized cost accounting systems. It has been used primarily in trials of two different ways of performing the same procedure, such as standard versus minimally invasive surgery. The detailed microcost study is done to establish the cost of the procedures of interest, while standard prices are used to measure the other costs of other components of care.

**ADDITIONAL DATA**

In some trials, it is feasible to collect economic data by accessing the computerized databases used by insurers or health care systems. For example, the Medicare program in the United States is billed by hospitals, physicians, and other health care providers for covered services. The Health Care Financing Administration has extensive computer files containing dates of service, the services provided, and the charges billed by the provider. However, these records do not contain bills for services not covered by Medicare, most notably outpatient drugs and nursing home stays. Most insurance companies have comparable computer records of billed services. Even health care systems that do not bill patients directly for services (e.g., the Department of Veterans Affairs, the Kaiser Permanente Health Maintenance Organization) have detailed administrative records of services provided to individual patients.

In several studies, clinical trial databases have been linked with an administrative database to obtain utilization and billing data for trial participants. An administrative database can provide the data needed for economic analyses that use either billed charges or resource utilization as the starting point of cost finding. While potentially a powerful research tool, administrative databases have several major limitations. Permission to access computerized data may be very difficult to obtain, especially given recent concerns about the privacy of medical records. Furthermore, since the United States does not have national health insurance, the needed administrative records are usually scattered among many different databases. Finally, databases constructed for administrative purposes may not be simple to adapt for research use. There is, for example, a delay of several years before Medicare data become available to researchers. Administrative data linkage is most feasible when a randomized trial is conducted in a single health care system that has comprehensive records; a Veterans Affairs administrative database, for example, can use the Veterans Affairs administrative databases (7).

**SCOPE OF DATA COLLECTION**

In economic analyses of medical therapies, it is useful to separate the costs of the intervention itself from the costs of complications that may be induced (or averted) as a result of its use. An intervention may “pay for itself” by reducing complications or disease progression sufficiently. Therefore, the scope of economic data collection needs to be sufficiently broad to capture costs of any complications of the treatment. For example, tissue-plasminogen activator (t-PA) for acute myocardial infarction may cause bleeding (increasing costs) but may reduce the extent of infarction and thus prevent heart failure (reducing costs). To provide a fair and complete comparison, a trial of t-PA needs to measure the cost of treating bleeding and heart failure. Data collection should continue long enough to document any plausible late benefits or complications of therapy.

The design of an economic analysis will have to consider the appropriate scope of data collection. Should a trial of heart failure treatments include the cost of treating unrelated...
conditions such as pneumonia or depression? The goal is to enhance the “signal” of disease treatment costs relative to the “noise” of unrelated medical costs. The issue is that some of the “unrelated diseases” may be adverse effects of treatment, and omitting their costs would bias the analysis. This trade-off is not unique to economic endpoints, since randomized trials often have to choose between disease-specific mortality or total mortality as the primary endpoint. Using disease-specific mortality increases the power of the trial but carries the risk of ignoring adverse effects of treatment that increase other causes of death. Reporting both total costs and disease-specific costs allows the reader to judge the economic outcomes more completely, analogous to reporting both total and disease-specific mortality.

There has been considerable controversy about whether to include nonmedical costs in the economic analyses, for example, out-of-pocket costs of treatment for a patient and his or her family—such as transportation to and from medical facilities—and the cost of care given at home by family members. There is a general consensus that these costs should be measured when they are likely to be large or fall disproportionately on one treatment arm. An economic comparison of inpatient versus outpatient surgery, for example, would provide very different results if the costs of convalescent care provided at home were not measured. There is less consensus on whether costs related to lost time from work should be included as an “indirect cost” of treatment in a clinical trial. Most studies report time lost from work as a secondary outcome measure but exclude it from the calculation of costs in the economic analysis.

**DATA ANALYSIS**

Analysis of economic data poses several technical difficulties (8). The distribution of costs tends to be skewed to the right, because most patients will use few resources and incur low costs, while a few patients will use many resources and incur high costs. (In any health insurance plan, three quarters of the costs arise from one quarter of the patients.) The pronounced skew of cost data reduces the power of the $t$ test and other techniques that assume a normal distribution. The simple alternative is to use methods that have less-restrictive assumptions about the data distribution, such as the nonparametric Wilcoxon rank-sum test. Another approach is to transform the cost data to a less-skewed distribution. The log transformation is particularly useful in this regard, because log-transformed costs can be used in regression models to test the association between clinical variables and the cost of care. To avoid taking the log of zero, a small constant, such as $\$1$, must be added to all costs before log transformation [$\log (0 + 1) = 0$].

A more difficult issue in data analysis emerges from accumulating costs over fairly long follow-up times. In most clinical trials, there is a lengthy period of patient accrual, and, unless all patients are followed to a fixed time (e.g., 3 years), there will be censored observations. Censored observations of mortality led to the development of actuarial methods, exemplified by the Kaplan-Meier survival curve, to accommodate variable lengths of follow-up. A similar method can be used to summarize cost data with variable follow-up times. Etzioni et al. (9) described a method by which follow-up is divided into discrete intervals $t_k$, and the mean costs of patients under observation in each interval are designated $F_t$. The mean cumulative cost per patient through time $t_k$ can then be estimated as

$$c(t_k) = \sum_{i=0}^{k} S(t_i) \times F_i,$$

where $S(t_i)$ is the proportion of patients alive at $t_i$. The variance of this cumulative cost can be estimated by bootstrap resampling of patients and recomputation of $C(t_k)$.

Studies that last longer than 1 year also need to consider inflation and discounting. Cost inflation is a familiar problem: $\$1$ in 2002 has less value than it had in 1995. When costs are measured over a prolonged follow-up interval, dollars need to be normalized to a single year to keep the monetary units constant over time. Current dollars can be converted to constant dollars by using the Consumer Price Index or Gross Domestic Product deflator. While the medical component of the Consumer Price Index might seem to be ideal for this purpose, it overstates the effects of inflation because it is also affected by the introduction of new medical technology and other changes in medical practice.

Discounting is a separate issue from inflation. Even in a stable, inflation-free economy, lenders are paid interest by borrowers. Thus, if the interest rate is 3 percent, $\$1.00$ today is worth $\$1.03$ in a year: this is the time preference for money. In long-term studies, the effect of inflation is handled first (based on calendar year), then the constant dollars should be discounted at a rate of roughly 3 percent per year after time zero, which is the date of randomization in a typical clinical trial.

International studies pose unique methodological issues for economic evaluation. Data collection in such studies is commonly limited to measures of resource consumption, since individual patients are not billed for services in many health care systems. The prices paid for resources vary considerably among countries, so the total cost of care should be computed by using alternative sets of price weights, one for each country in the study. However, it is controversial whether the economic analyses should be limited to the subset of patients enrolled in a particular country or whether the alternative price weights should be applied to the resource consumption profiles found in the entire trial population. Since there is a great loss of statistical power in subsets, I prefer the method of applying alternative country-specific price weights to the entire trial population.

**COST-EFFECTIVENESS ANALYSIS**

Cost is only one factor in the decision concerning whether a new therapy should be adopted. Therapies that yield better clinical results at a lower cost should obviously be adopted, but, more commonly, the new therapy provides better clinical results at a higher cost. In this situation, the extra cost must be weighed against the added benefits to judge the...
value of the new therapy. Policy analysts have developed a formal framework to assist in these judgments, termed cost-effectiveness analysis (10, 11).

The preferred measures of effectiveness in a cost-effectiveness analysis are the life-year (LY), which values reduced mortality, or the quality-adjusted life-year (QALY), which values reduced morbidity and improved quality of life as well as reduced mortality (3). These measures are patient-centered outcomes, emphasizing that improved health is the bottom line for any therapy. Surrogate endpoints such as blood pressure or cholesterol levels generally are not used in cost-effectiveness studies because they are not directly perceived by patients.

Cost-effectiveness analysis compares a new treatment with the next best alternative. Therapies are “cost-effective” only in comparison with the other options, reflecting the reality that there are choices for the patient and that money spent on one treatment is not available for use elsewhere in the health care system. Thus, the cost-effectiveness (CE) ratio is defined as follows: $$CE = \frac{(cost_i - cost_j)}{(life-years_i - life-years_j)}$$, where cost, is the total cost of using treatment i, and life-years, is the patient’s number of expected years of survival when given the therapy.

The cost-effectiveness ratio can be computed by using data from a clinical trial. Cost, can be measured as the cumulative cost per patient by using the modification of the Kaplan-Meier method described in the previous section of this review. Life-years, can be measured as the area under the standard Kaplan-Meier survival curve. For consistency, the numerator and denominator of the cost-effectiveness ratio are discounted at the same annual rate, typically 3 percent per year (3). While discounting future dollars seems quite natural and appropriate, discounting future survival benefits of therapy has been controversial, but this principle is well accepted by medical economists, since discounting costs without discounting benefits can lead to nonsensical policy recommendations (12). Discounting benefits also recognizes that patients value immediate outcomes more highly than future ones, a fact well known to those attempting to promote healthy lifestyles to prevent disease.

The cost-effectiveness ratio estimated in a clinical trial has both statistical uncertainty and variability because of parameters of the underlying model, and each should be assessed. Several methods have been proposed for gauging the statistical variability of the measured cost-effectiveness ratio (13–15). Bootstrap resampling of the patient population is arguably the best method because it makes no assumptions about the distribution of the data. In each bootstrap sample, the cumulative costs and life-years of survival are recalculated, and the cost-effectiveness ratio is recomputed. The results of multiple bootstrap replications can be displayed graphically (difference in cost between treatment 1 and treatment 2 on the vertical axis; difference in life-years on the horizontal axis). The percentage of bootstrap replications below any cost-effectiveness threshold (e.g., below $50,000 per life-year added) can be readily measured from the graphic display. Cost-effectiveness models should also be subject to a “sensitivity analysis” in which key parameters, such as the discount rate or source of cost weights, are varied over plausible ranges to assess whether the results are sensitive to assumptions in the model.

There is no absolute standard for an acceptable cost-effectiveness ratio. The most common benchmark is provided by renal dialysis, which costs roughly $35,000 per year to provide and to keep one patient alive for 1 year, albeit at a reduced quality of life. Cost-effectiveness ratios of less than $35,000–$50,000 per life-year added are considered acceptable in the United States, since this is roughly the cost-effectiveness of renal dialysis. Cost-effectiveness ratios of $50,000–$100,000 per life-year added are more arguable but are also generally accepted in the United States. Therapies that cost more than $100,000 per life-year are generally “economically unattractive,” even in the United States.

Cost-effectiveness analysis was developed for use in policy analysis and decision models, and it has several limitations when applied to the primary data from a clinical trial. The biggest is that follow-up in a clinical trial ends after a few months or years, and the full effectiveness of a therapy in terms of life-years added is not observed completely. This is a particular issue for drugs, devices, and procedures that reduce mortality but have very high initial one-time costs. A limited follow-up will capture the complete cost of therapy but only part of its benefit, introducing a bias into the cost-effectiveness estimate. It is feasible to make assumptions about future costs and survival and to use a model to extrapolate the trial results and compute a long-term cost-effectiveness ratio (16). Current methodological research in medical economics includes developing standard, consensus methods that bridge the divide between the demands of current cost-effectiveness methods for a lifetime perspective and the realities of clinical trial results that provide a limited follow-up.

**SUMMARY**

Economic endpoints are increasingly common in randomized clinical trials. As a relatively new addition to the field, methods of measuring and analyzing cost data are still evolving. These developments will be stimulated by the demands of the public for efficient and effective medical care, which will be based on the findings of clinical trials.

**REFERENCES**

1172–7.