Experimental versus Observational Data in the Economic Evaluation of Pharmaceuticals

MICHAEL F. DRUMMOND, DPhil

It is commonly suggested that there are two competing approaches to the economic evaluation of health technologies. In a trial-based study, economic data (e.g., use of resources) are collected concurrently with the clinical study. In a modeling (or integrative) study, data from a number of sources are synthesized.

In the first case, the economic data can be viewed as experimental and are typically analyzed in the same way as the clinical data. In the second case, the data for the economic evaluation could be either experimental (e.g., efficacy data from clinical trials) or observational (e.g., resource-use data extracted from patient chart review or a claims database). Indeed, there are various potential sources of observational data in economic evaluations, including claims databases, epidemiologic studies such as the Framingham Heart Study, and surveys of patients’ health-state-preference values. Even trial-based studies use some data obtained from outside the clinical trial, such as the prices (or unit costs) of health care resources.¹

Recently there has been considerable debate about the relative merits and demerits of trial-based and modeling approaches to economic evaluation. In an editorial in the New England Journal of Medicine, Kassirer and Angel² argued that “bias can compromise even original scientific studies, but we believe that the opportunities for introducing bias into economic studies are far greater, given the discretionary nature of model building and data selection in those analyses.”

Furthermore, in its draft Principles for the Review of Pharmaco-economic Promotion,³ the Food and Drug Administration (FDA) argued that research to substantiate pharmaco-economic claims, including cost-effectiveness and quality-of-life claims, must meet traditional standards for adequate and well-controlled studies. It also suggested that models to provide estimates of pharmaco-economic parameters be used only when it is impractical or impossible to gather data using adequate and well-controlled studies. Whereas the FDA’s principal concern is with the effectiveness data used in pharmaco-economic studies, it clearly views modeling approaches as representing weaker methodology.

Finally, in the revision of its guidelines for economic evaluation of pharmaceuticals, the Commonwealth of Australia⁴ recommended that a preliminary cost-effectiveness assessment based on data from the most relevant clinical trials be made prior to any extrapolation(s) and/or assumption(s) being used to transfer or generalize the evidence. In making this recommendation, it argued that “trial-based comparisons are the most internally valid and pay due heed to biostatistical and epidemiological rules.”

Rittenhouse and O’Brien⁵ have characterized the problem as one of a tradeoff between internal validity and external validity. Trial-based economic studies have high internal validity, in that the differences between treatment groups are unlikely to be biased since a randomized experimental design is employed. However, they may have relatively low external validity, in that randomized controlled trials (RCTs) are often undertaken under conditions atypical of normal practice⁶ and therefore the results may not be generalizable.

On the other hand, a modeling approach offers the potential for generalization from trial to practice, and for transferring the results to other locations, in that data relevant to any given setting can be included in the analysis. However, such studies run the risk of low internal validity, in that if observational data are used, these may contain biases that the analyst is unaware of.

One possible solution would be to undertake clinical trials of pragmatic design, incorporating economic data capture. Pragmatic clinical trials are randomized as normal, but are undertaken in settings that approximate regular clinical practice, have few clinical procedures mandated by the protocol, seek to enroll normal caseloads, have the sample sizes set to accommodate socioeconomic variability, and follow patients for long periods. However, truly pragmatic trials can be costly to undertake and are unlikely to be undertaken in every conceivable setting.

Address correspondence and reprint requests to Professor Drummond: Centre for Health Economics, University of York, Heslington, York, YO1 5DD, United Kingdom.
Therefore, a balance needs to be struck between the use of experimental and observational data in economic evaluations. This paper explores this issue in the following manner. First, in the next section the data requirements for economic evaluation are outlined, explaining that data can come from both experimental and observational sources. Then the advantages and disadvantages of using various types of data are explored. Finally, a few conclusions are drawn.

Data Requirements for Economic Evaluation

Whether economic evaluations are trial-based or modeling studies, the data requirements are the same. First, data defining comparative treatment effectiveness are required. Often these will come from controlled clinical trials and can be used directly in the economic study. However, on some occasions, adjustments may be necessary if the trials concerned were undertaken under artificial conditions. For example, in their study of prophylaxis against deep-vein thrombosis (DVT) after total hip replacement, O’Brien et al. adjusted the rates of DVT detected by venography in the trial to allow for the fact that some DVT would be clinically unimportant and would not have come to the notice of the patient or the physician in regular clinical practice. A similar argument applies to ulcers detected by endoscopy.

On the other hand, some economic evaluations project long-term outcomes (e.g., survival from intermediate endpoints measured in trials (e.g., serum cholesterol) by using epidemiologic models. Here the outcome data are clearly not taken directly from the trial and depend on the validity of the model.

Second, data describing the physical quantities of resources consumed by the treatments being compared are required. For example, these include amounts of time expended in medical and nursing care, drugs, hospitalization, community services and the patient’s family’s own resources, in traveling to health care facilities or in providing informal nursing support in the home. These data can either be collected alongside a trial or be obtained from a range of other sources, including routine statistics, local surveys, and expert opinion.

Third, data delineating the unit costs or prices of resources are required. (These enable us to calculate the total and incremental costs of one therapy over another.) Often these data come from routine data collection or observational studies, although they could in principle be collected alongside the trial.

Fourth, data showing how therapies affect patients’ qualities of life are required. Since these rely on direct reports from the patients experiencing the health state(s) concerned, they are usually gathered alongside the trial.

Finally, data that describe the preference individuals place on states of health may be required, especially if the aim of the economic study is to calculate the cost per quality-adjusted life year gained. Such data can be obtained from the trial, by asking patients to value their own current health states by standard-gamble or time-tradeoff methods, or from surveys of members of the general public, who may be presented with series of health-state scenarios.

Advantages and Disadvantages of Experimental versus Observational Data

If many of the data items used in economic evaluations can come from either experimental or observational sources, which is better? A primary message of this article is that this question cannot be answered in the aggregate. It needs to be answered in respect of each category of data.

At the outset, it is worthwhile recognizing that data do not fall neatly into categories. For example,
figure 1 outlines the data that are typically collected alongside a clinical trial. Few are unambiguously "clinical" or "economic." Therefore, it is meaningless to say that the "clinical" data in an economic study should be taken from experimental studies and that the "economic" data from observational studies. It is important to be more specific.

Sometimes, primary elements of clinical data could be viewed as "economic." For example, the EPIC trial, which assessed the efficacy of monoclonal antibody in high-risk coronary angioplasty, used a composite primary endpoint of death, non-fatal myocardial infarction (MI), unplanned surgical revascularization, unplanned repeat percutaneous transluminal coronary angioplasty, unplanned implantation of a coronary stent, or insertion of an intra-aortic balloon pump for refractory ischemia.

The motivations for selection of such an endpoint are twofold. First, all the components of the endpoint are indicators of coronary insufficiency. Second, some of the events (such as death and MI) are quite rare, and a trial designed to detect differences in these alone would need to recruit a large number of patients in order to have sufficient statistical power.

However, the selection of such a composite endpoint introduces a complication, in that the response being measured is no longer solely biological, but likely to be influenced by a number of economic and organizational factors peculiar to the setting or the health care system. For example, some settings may have less resources to provide access to the procedures concerned. A number of the components of the endpoint will also be used for costing the therapies concerned.

### DATA ON EFFECTIVENESS

In general, the data on clinical effectiveness used in economic evaluations should be taken from RCTs, because the inherent biases in data from uncontrolled or observational studies represent a major threat to validity.

However, there are a number of concerns about the data from clinical trials that might cause us to depart from this general principle.

First, many of the protocol-defined clinical events in clinical trials may not be clinically relevant. For example, it is known that much of the DVTs detected in clinical trials by venography would not have become clinically important (i.e., producing symptoms or leading to pulmonary embolism). Similarly, protocol-defined MIs may include mild infarcts that would require nothing but close observation. Therefore, at least the clinical-event data would need to be classified in relation to their clinical relevance or adjusted before use in an economic evaluation—biases could be introduced if the rates of events between therapies differed in relation to clinical importance, and hence cost. O'Brien et al. found this in their analysis of clinical trials of warfarin and low-molecular-weight heparin in prophylaxis against DVT. Much of the superior efficacy of low-molecular-weight heparin was in the prevention of distal DVT, which is not as clinically important as proximal DVT.

Second, the clinical endpoints measured in trials may not, of themselves, be suitable for use as the measure of benefit in the economic evaluation. For example, the endpoints may be intermediate outcomes, such as percentage cholesterol reduction. Although there may be some interest in comparing cholesterol-lowering therapies in terms of their costs per percentage cholesterol reduction, the more relevant economic comparisons are in terms of cost per life year gained or cost per quality-adjusted life year gained, since this facilitates decisions about the alternative uses of health care resources.

Furthermore, even if the trial measures final endpoints such as survival, there may be a need to extrapolate beyond the trial observation period. For example, Mark et al. compared tPA and streptokinase in terms of their costs per life year gained. Although the trial measured survival at one year, calculation of the life years gained required knowing the likely survival of patients alive at the end of the trial observation period. Therefore, the authors used observational data from a clinical database of patients experiencing nonfatal MI to extrapolate from one to 14 years and then a statistical extrapolation for 14 years and beyond.

Other secondary clinical events, such as some side effects, may be rare, and the trial may not be powered to detect differences in their occurrences for the interventions being compared. Indeed, some events may be so rare that none occur during the trial. Since adverse events can have economic consequences, in resource use and reductions in quality of life, it may be important to estimate the probabilities of their occurrences for the economic evaluation. Therefore, it may be necessary to undertake case-control studies using observational data, although clearly such studies have an inferior design and run the risk of introduction of biases.

In addition, if only a few events occur during the trial, it may be difficult to obtain a precise estimate of their economic consequences in resource use. Therefore, it may be necessary to use observational data to supplement or adjust those obtained during the trial.

Thus, it is difficult to give clear guidance on whether data on secondary clinical events, such as side effects, should be experimental or observational. In an ideal world they would be experi-
mental, but this may not always be possible. Given the biases potentially introduced through the use of observational data, it would be important to assess (through sensitivity analysis) the quantitative importance of the method of estimating the economic consequences of side effects, for the overall result of the economic evaluation.

RESOURCE QUANTITIES

Resource quantities can be estimated either concurrently with a clinical trial (i.e., experimental data) or in parallel studies (i.e., observational data). Insofar as the weight of arguments is in favor of trial-based estimates of clinical effectiveness (see above) and parallel studies for estimation of unit costs or prices (see below), the debate over the estimation of resource quantities is the cornerstone of the broader debate about experimental versus observational data in economic evaluation.

The arguments for estimation of resource quantities inside or outside clinical trials are both evenly balanced and complex. In favor of within-trial estimation, one could argue that, first, many of the resource drivers (e.g., hospital admissions) are closely linked to the clinical endpoints; second, there is therefore no strong reason to treat these therapeutic responses differently from those more typically characterized as clinical responses. (An example is given above of one major clinical trial where the clinical and economic endpoints were indistinguishable.) These arguments would appear to underpin the FDA’s position on promotional claims for pharmaceuticals. In short, if one wants to claim a given impact of an intervention, why not set up a study to demonstrate the impact within an experimental setting? Economic studies that impute the given impact, by making a number of assumptions, are inherently weaker.

However, the arguments against estimating resource quantities alongside the trial are also quite strong. First, we know that many trials are undertaken in fairly artificial settings, so the resource use observed in the trial may not reflect that likely to be observed in regular clinical practice. The “artificiality” of trials has been well documented.” In the main, this results from the circumstances that many trials are undertaken exclusively in specialist centers, that certain elements of resource consumption (e.g., visits and tests) are driven by the trial protocol, that the patient and the physician are “blinded,” and that the closer monitoring of the patients in the trial means that key clinical events are often detected and dealt with sooner than in real-life situations.

Second, the resource consumption observed in clinical trials relates not only to the biological responses of the patients to the therapies administered but also to the responses of the health care systems in existence in the settings studied. Therefore, a given clinical event (e.g., incidence of suspected MI) may provoke different resource responses depending on where the event occurs. For example, availability of out-of-hours cover may influence the patient’s ability to see a physician (either in ambulatory care or in the emergency room). Different diagnostic facilities could influence the determination of a differential diagnosis, and hospital formulary policies and availability of beds could influence the likelihood of admission and the therapy given.

Whereas in a clinical trial with balanced randomization these factors will not affect the relative risk reduction (or odds ratio) used to compare therapies; they could affect the absolute size of the economic (resource) effect. Also, in multicenter trials they could greatly complicate the economic analysis.

Therefore, the arguments for excluding resource capture alongside clinical trials are based on the proposition that the primary purpose of the trial is to estimate the patient’s biological response to the therapy administered. This is usually assumed, and often shown, to be independent of the setting in which the therapy is delivered. Estimation of the economic response to therapy is inevitably confounded within clinical trials unless patients are randomized to setting as well as to therapy. Where does this leave us with respect to estimation of resource use alongside clinical trials? There appear to be two guiding principles. First, resource data capture is more likely to be relevant if the trial is of pragmatic design. Second, resource impacts are likely to be less confounded, and more transferable, in situations where the health care system responses to a given biological response do not vary greatly from setting to setting. For example, in a trial of a drug to reduce the incidence of subsequent MIs in patients ‘at risk of heart attack, the least confounded data are likely to be those relating to the incidence of MI. The data on the rates of hospital admission for MI are likely to be influenced in part by the setting, but still may give a reasonable estimate of the value of a drug that has this protective effect.

Data on the numbers of hospital days of and procedures used for patients admitted for MI are likely to be greatly influenced by the settings in which the trial is conducted and therefore may be both confounded and difficult to transfer. However, such data will be highly relevant for the setting in which the trial took place, and for similar settings.

One interesting issue is whether the demonstration of reduction in resource use in a given trial setting (however atypical that is) adds to the credi-
bility of the economic case. For example, if drug X reduced hospitalization in an HMO in California, would this greatly impress health care decision makers on the East Coast of the United States, or in other countries? In a sense, this is all that a clinical trial can accomplish for the clinical endpoints. That is, generalizations from a clinical trial are valid only for the same therapy applied with the same amount of skill to similar patients in settings with similar supportive care, in which circumstances one would expect similar outcomes. One cannot conclude that the trial’s results would necessarily obtain in one’s own setting.

Perhaps for economic studies the dichotomy between what was observed in a clinical trial and what could be achieved in one’s own setting is a little more obvious. However, it is there for the clinical data nevertheless. Therefore, for a decision maker interested in economic outcomes, the FDA argument that the effect should be observed in adequate and well-controlled studies probably carries less force. The decision maker is more likely to be influenced by actual events in his or her own setting.

In summary, because of the complexity of the arguments presented above, it is difficult to give firm guidance about the choice of a balance between experimental and observational data for the estimation of resource quantities. In general, one might argue that the resource-precipitating events (e.g., incidence of, or hospitalization for, MI) should probably be estimated within the clinical trial, whereas the resource consequences of the precipitating events (e.g., days of hospitalization for MI or number of revascularization procedures following MI) should probably be estimated from observational data in the settings for which economic data are required.

However, this approach could be misleading if the interventions being evaluated differed not only in their impacts on the frequency of events precipitating hospitalization but also their intensity (and hence resource use). The extent of this problem would be difficult to assess a priori, but its effect could be minimized if the events were carefully graded in terms of clinical severity (e.g., severity of MI).

PRICES (OR UNIT COSTS) OF RESOURCES

It is well known that both relative and absolute prices vary from setting to setting. Therefore, it is highly unlikely the costs that result from applying a given set of prices to the resource-use data will be transferable. Most users of economic evaluations are likely to want their own prices, or those from similar settings, included in the study.

In addition, the price data observed during a given clinical trial will reflect the trial recruitment, not real life. For example, in a multicenter trial, total cost could be inflated by recruitment in a single, large, expensive center. If the randomization is balanced by center, this will not confound the cost comparison between the two or more therapies considered in the trial. However, it could affect the incremental cost-effectiveness ratio, which is problematic, since it is common practice to compare the incremental ratio obtained in a given study with those obtained previously for other health care interventions.14

Therefore, the balance of argument is in favor of obtaining prices from observational studies outside the clinical trial. Normally they are obtained from published statistics (e.g., physician fee schedules) or from local studies.

There appears to be only one strong argument for using the prices observed in the trial. That is, it could be argued that the relative prices of resources influence the combination of resources in the provision of therapy, which in turn influences outcome (see figure 2). Since the outcome of treatment is used in the denominator of the cost-effectiveness ratio, it is only logical to use, in the numerator, the actual economic variables that produced the outcome.

However, the argument breaks down if one assumes that, in adopting a new therapy, one can mandate the mixture of resources used. Therefore, the decision maker is interested first in the resource mix used to produce the given outcome and the costs, in his or her setting, of the chosen resource mix.

DESCRIPTIVE QUALITY-OF-LIFE DATA

The arguments relating to the collection of descriptive quality-of-life data are similar to those for the clinical endpoints. That is, in general they should be collected in the trial, unless the trial is being undertaken under highly artificial conditions. For example, there may be cases where protocol-mandated investigations or follow-up visits that would not take place in regular practice have a major impact on patients’ quality of life. However, at most this may suggest that the data should be adjusted, not that observational data should be used instead.

HEALTH-STATE-PREFERENCE VALUES

In principle, health-state-preference values could be obtained either from the patients enrolled in the trial, by asking them to do standard-gamble or time-tradeoff assessments of their “current state,” or from a survey of individuals outside the trial. In the latter case, various relevant health-state scenarios
Experimental vs Observational Data

**Discussion and Conclusions**

It should be clear from the discussion above that there is no overall (aggregate) conclusion about the relative merits and demerits of experimental versus observational data for economic evaluation. The choice of analytic approach relates to the category of data concerned. Clinical and descriptive quality-of-life data should probably be trial-based. Unit costs or prices should probably be obtained from sources outside the experimental setting. The question of whether resource quantities should be obtained from within, or outside, clinical trials is probably the most vexing: the answer probably depends on where the clinical trial is on the spectrum between being explanatory or pragmatic.

Thus, it is all a question of balance. Figure 2 summarizes the arguments presented here. These conclusions are inevitably tentative, and much more discussion is required. However, the conclusion is that some categories of data in economic evaluations should definitely be obtained in RCTs, and others should definitely be obtained in observational studies. However, many important categories of data fall in the methodologic “grey area” indicated on figure 2.

A practical consequence of the arguments outlined here is that the FDA’s requirement of two “adequate and well-controlled studies” is not as daunting as it might seem at first sight. That is, the efficacy and quality-of-life data used in an economic evaluation should be gathered alongside a controlled clinical trial, although on occasions they may require adjustment to reflect regular clinical practice. However, some of the other data, such as unit costs or prices, may be better gathered in observational studies or from routine data sources.

The key judgments relate to the method for gathering data defining resource quantities and health-state-preference values. These may require a hybrid approach, with some data collected alongside clinical trials and some gathered from observational studies. In the case of health-state-preference values, an instrument such as the health utilities index (HUI) or the EuroQoL-EQ5D could be administered in the trial instead of undertaking standard-gamble or time-tradeoff measurements directly on patients.
In the case of resource quantities, the data collected in the trial could be limited to the recording of major cost drivers (e.g., hospitalizations) rather than collecting all the details of resource use. The detail may in any case be influenced by the setting of the trial, which would limit transferability. If those expressing concerns about the validity of pharmacoeconomic studies were to accept this approach, the data-collection burden for those wishing to base studies on RCTs would not be greatly increased, as many of these trials already include collection of data on quality of life and major resource items such as hospitalizations.

The choice of experimental or observational data represents a series of compromises, between internal validity and external validity, between specific accuracy or broader policy relevance, and between immediate solutions and long-term study. If there is one guiding principle in making this choice, it is the following. We should not generally use observational data to establish or attribute a difference between therapies, but they can be used to estimate the economic consequences of such a difference.

An earlier version of this paper was presented at the CePOR ’98 conference, University of North Carolina, 31 March-2 April, 1998, where many useful comments were received. The author is also grateful to James Mason, Bernie O’Brien, Trevor Sheldon, and an anonymous referee for comments on an earlier draft and to Vanessa Windass for secretarial assistance.

References