The Value of Observational Datasets in the Economic Evaluation of Pharmaceuticals

Keith Tolley[1], Francis Pang[2] , Martin Brown[1], Clare Hibbert[3], Eddy Beck[4]

[1] Outcomes Research, Pfizer Central Research, Sandwich, Kent, UK
[2] Centre for Health Economics, University of York, York, UK
[3] MERCS, Intensive Care Unit, Royal Hallamshire Hospital, Sheffield, UK
[4] Imperial College of Medicine, London

Keywords: Economic evaluation; methodology; observational

Paper presented to the Health Economists Study Group Winter Meeting, University of Newcastle, January 2000

Work in progress: Please do not quote

---

1. Introduction

A well-known definition of economic evaluation is a comparative analysis in terms of its costs and consequences (Jonsson, 1993). Despite this seemingly standard definition and some agreement about the fundamentals of cost-effectiveness/cost-benefit analyses, no similar standard design framework for economic evaluation exists and there is much disagreement about specific details (Sloan, 1995). In many cases, researchers conducting an economic evaluation make use of whatever data is available. This has led to concerns of bias (Hillman et al., 1991) and an inherent danger of constructing a "Frankenstein's monster" (O'Brien et al., 1996).

Economic evaluations critically rely on the assessment of clinical effectiveness. The data can come from a single study, a systematic overview of several studies or an ad hoc synthesis of several sources. So any limitations which weaken the assessment of effectiveness weaken any economic evaluation based on it. Economic evaluation can be based on either prospective or retrospective frameworks (Adams et al. 1992, Drummond et al., 1991)

In terms of the single study, the gold standard for assessing efficacy of interventions is the randomised, double blind controlled trial. This design has the highest internal validity- i.e. freedom from bias - and sits at the top of a hierarchy of clinical study designs which has been proposed in the clinical epidemiology literature, followed by cohort/longitudinal studies, case-control studies and case-series (Sackett et al., 1985). The gold standard for evaluating effectiveness (i.e. does the drug achieve additional benefits in real clinical practice) is widely recognised as the pragmatic RCT. Pragmatic clinical trials are randomized as normal, are undertaken in settings that approximate regular clinical practice, have fewer clinical procedures mandated by the protocol, seek to enroll normal caseloads, have the sample size set to accommodate socioeconomic variability and follow patients for long periods. However only a relatively small

proportion of clinical studies are randomised controlled trials (most are conducted by drug companies), fewer still are pragmatic and few of either type are used for full economic evaluations of drugs. If economists ignore the other types of study designs, then there is a potential (depending on your viewpoint) for a significant loss of wealth of information that could be used to construct cost-effectiveness arguments. Nixon and Pang (forthcoming, 2000) found that economic evaluations conducted in Japan were predominantly based on case-series or cohort designs, possibly as a consequence of the lack of randomised controlled trials performed in the country.

Case reports, case-series, secular trends, case-control studies, and cohort studies are collectively embraced under the umbrella of observational studies or non-experimental study designs. In these studies, the investigator does not control the administration of the therapy, but rather observes and evaluates the results. However, in this paper we focus on the use of longitudinal/cohort designs containing patient-specific data and the value of such datasets, relative to the construction or use of RCT study designs, for economic evaluations of pharmaceuticals. We wish to explore and discuss whether the use of such observational data is an acceptable alternative to RCT data in the economic evaluation of drugs currently being prescribed, or those still in development for which drug companies desire third party payer reimbursement status.

In this paper we discuss the data requirements for full economic evaluations of pharmaceuticals (i.e. covering CEAs and CUAs), and how observational datasets could potentially be used to partially or fully supply that data. We then present an initial review of literature from the HEED database for 1995-99 of the use of observational or RCT data in applied pharmacoeconomic evaluations. Subsequently, we discuss issues relating to the relative value of such study designs for conducting such economic evaluations, and then suggest some multivariate techniques for aiding economic analysis using longitudinal or cohort data. Finally, a number of longitudinal datasets containing resource use and outcomes data have been developed in recent years in the UK. These datasets are briefly reviewed. Some conclusions are presented regarding the potential for, and value of, pharmacoeconomic evaluations using non-RCT data.

2. Data Requirements for Pharmacoeconomic Evaluations

The majority of published pharmacoeconomic studies take the form of cost-effectiveness analyses (CEAs) or, increasingly, cost-utility analyses (CUAs), representing 80.6% of all studies (Elixhauser, 1998). Pharmaceutical companies commission many of these studies often as phase IV evaluations for drugs that have already been launched on the market. Other pharmacoeconomic studies are conducted to support the launch of newly licensed drugs. The development in a country of a fourth hurdle, for example through the Pharmaceutical Benefits Advisory Committee (PBAC) in Australia (Commonwealth Department of Human Services and Health, 1995) and the National Institute for Clinical Excellence (NICE) in the UK (NICE, 1999), may increase the number of pharmacoeconomic studies published at about the same time as the efficacy results for new compounds. These pharmacoeconomic studies are more likely to have made use of data from phase III randomised controlled trials, whereas phase IV studies tend to employ observational study designs or, less

likely, pragmatic RCTs. However, observational designs or pragmatic controlled trials might also be used for special Outcomes Research studies conducted alongside phase III clinical trials in order to improve the external validity of the results relative to internal validity. If economic or outcomes data have not been collected during phase III, or time and resource constraints limit the possibility for primary data collection in phase IV, then decision modelling designs incorporating data from meta-analyses/systematic reviews, large retrospective observational datasets or expert opinion (e.g. as in Delphi Panels) are frequently employed. Table 1 presents a summary of the drug development stages for demonstrating safety and efficacy of a new compound.

Three sets of core data are required for full CEAs or CUAs of pharmaceuticals. These are:

- resource use consequences: e.g. hospitalisation, primary care utilisation, drug consumption, procedures, tests, consultations.

- the probability of clinical events and outcomes: e.g. treatment outcome, adverse events, recurrence.

- disease/health outcomes e.g. cure, health improvement, survival, QoL, utilities.

A pharmacoeconomic study may use data for all of these from one source only, such as an RCT or a prospective observational study design. However, many studies use different sources for each core data requirement in an economic evaluation. All designs should obtain basic data on the study population characteristics such as age/sex, concomitant conditions Observational study designs could therefore be used for two broad purposes in CEAs/CUAs:

a) Providing partial data; e.g. for resource use only
b) Providing full data for resource use, probability of clinical events and outcomes and disease/health outcomes.

Drummond (1998) proposes that the choice of experimental versus observational data for economic evaluation is related to the category of data concerned. Drummond suggests that clinical and descriptive quality of life data should probably be trial-based, whereas unit costs or prices should probably be obtained from observational sources. However he concedes that the issue whether resource quantities should be obtained from within or outside clinical trials is probably the most vexing and depends where the clinical trial is on the spectrum between being explanatory or pragmatic (Schwartz, 1967).

The use of observational data to assign (unit) costs on resource use has been advocated to substitute cost measurement alongside RCTs (Gold et al., 1996; Sheldon, 1996) and has been applied to an economic evaluation of thrombolytic therapy (Kalisch, 1995). It is essential as a prerequisite to consider carefully as to whether the characteristics of the patients involved are comparable to those enrolled in the RCT.

Longitudinal/cohort datasets have frequently been recommended for other types of economic analysis, such as costing or cost of illness studies (CMR, 1997). If the health care resource use of newly diagnosed or treated patients with particular diseases can be tracked for a sufficient follow-up period, then an estimate of the incidence cost of disease can be derived. Incidence based costing studies can be used to estimate the economic benefits of a preventive drug or intervention, or for profiling variations in costs of treatment and care over time and different stages of disease. For example, the costs of schizophrenia or stroke are highest in the first 6 months after diagnosis, whereas the costs of HIV/AIDS tend to be greatest in the last 6 months. Cost studies based on longitudinal datasets are also useful for constructing iso-resource groups within a DRG type system. For example, the Department of Health in the UK have been assessing the development of ICU HRGs using data from a database that has been set up in a Sheffield hospital (Royal Hallamshire Hospital) to produce detailed patient specific estimates of ICU costs for different conditions and treatments (Edbrooke, 1999).

Table 1: Stages of Drug Development

Discovery	Research planning, organic chemistry, synthesis, molecular biology, basic screening
	↓
Pre-clinical	Pharmacokinetics, animal model determination, acute toxicity testing, mutagenicity/ carcinogenicity tests, dosage studies, main effect/ side effecrts
	↓
Phase I	First administration to man. Pharmacokinetic studies, healthy volunteer studies, tolerability studies.
	↓
Phase IIa	First studies in target patient population. Efficacy studies, toxicity studies, dosage studies, kinetics studies, supplementary animal studies.
	↓
Phase IIb	Extension of IIa with larger numbers of patients, perhaps widening the entry criteria
	↓
Phase III	Larger patient population studies, pivotal efficacy studies, case history data, adverse drug reaction data, final dosage studies, drug interaction studies, population pharmacokinetic studies
	↓
	Market Launch
	↓
Phase IV	Post-marketing surveillance studies, clinical experience data , patient acceptability data, spontaneous reports, comparisons with standard drugs and competitors.

3. Observational Dataset Designs

There are three main types of observational datasets that could be used to supply partial or full data in an economic evaluation:

a) Retrospective longitudinal/cohort datasets: these are datasets that already exist and contain data to various levels of detail on one or more of the core data areas outlined in the previous section. These datasets could be general collecting a core set of data for different diseases, or disease/condition specific. Data is collected for a defined time period (which may be a short time period, or over several years).

b) Prospective datasets: these are datasets which involve continual and ongoing monitoring and update. Otherwise they are the same as those in (a). Existing UK datasets such as GPRD and Memo contain data for different conditions for primary and/or secondary care, whereas specific prospective datasets exist for collection of ICU data (Edbrooke 1999) and HIV/AIDS (Beck et al, 1998) in England

c) Research question specific datasets: these are study designs developed for a specific research question (i.e. to assess the costs and effects of treatment and care for a condition or disease) with data collected for a defined time period.

4. Review of Study Designs in Applied Studies

As of February, 1998, 39% of economic evaluations on the Health Economic Evaluations Database (HEED) Database [note: this is not an observational database, but a database of bibliographic references and references which have been reviewed according to a standard format by health economists] used only primary observational data, defined as where observational data were the only data source recorded for probability of main clinical events, resource use and outcomes (Pritchard, 1998). This proportion had increased from 24% in 1992. Interestingly this growth in the use of observational data has been associated with the increase in number of cost-consequence analyses, which over half were based on observational data. As of December, 1999 the proportion of economic evaluations employing observational data is 49% on the NHS Economic Evaluation Database (NHS Centre for Reviews and Dissemination, 1996).

A new search of the HEED database conducted by the authors of this paper was conducted, with the search limited to:

- Pharmacoeconomic evaluations (i.e. a drug as the main intervention being evaluated).
- CEAs and CUAs onl;y.
- Studies identified in the database as applicable to the UK (and possibly other countries).
- Covering the period 1990-99.

A total of 71 studies were identified, 59 CEAs only, 6 CUAs only and 6 where both CEAs and CUAs were conducted. Table 1 presents the type of data used in these studies for the three core data items (as defined in the HEED review summaries). RCT data represented the majority source for all three items, in particular for clinical event and disease outcome data, as might be expected. Most of these studies were not pragmatic RCT designs. Observational data sources were more frequently used to provide resource use consequence data. In a few studies observational and/or RCT data provided the data for all three core items. Only a few studies using observational sources (less than 10) clearly use longitudinal or cohort data. Those that do generally use this data source for all three core items – mostly through a prospective design collecting primary data for the disease and drug therapy being evaluated. None of the reviewed studies made use of any existing prospective or retrospective database**.**

Table 2 : Number of studies using different types of data (review of 71 UK studies)

	Data source			
Core data	Observational	RCT	Observational & RCT	Other
Resource use consequences	21	27	3	20
Clinical events	19	37	6	9
Disease outcomes	19	33	6	13
All 3 sets of core data	9	14	1	6

5. Issues in the use of Longitudinal Observational Data (versus RCT)

The basic function of most observational research designs (see table 3) is to permit a fair, unbiased contrast to be made between a group with and a group without a risk factor or intervention. A good research design should perform the following functions:

(1) enable a comparison of a variable between two or more groups at one point in time, or in some cases, between one group before and after receiving an intervention or being exposed to a risk factor.

(2) allow the contrast to be quantified either in absolute terms or in relative terms

(3) permit the investigators to determine when the risk factor and the disease occurred, in order to determine the temporal sequence

(4) minimize biases, confounding and other problems that would complicate interpretation of the data.

Table 3: Relative merits of alternative epidemiological study designs

Study Design	Advantages	Disadvantages
Case Reports	Inexpensive and simple method for generating hypotheses	Cannot be used for hypothesis testing
Case Series	Simple quantification of incidence	Cannot be used for hypothesis testing because of absence of control group
Analysis of secular trends	Can provide rapid answers	Prone to confounding
Case-control study	Can study multiple exposures Can study uncommon diseases Simple and fast Less expensive	Selection of control problematic Possibly biased exposure data
Cohort/longitudinal study	Can study multiple outcomes Can study uncommon exposures Selection bias less likely Unbiased exposure data Incidence data available	Possibly biased outcome data More expensive If done prospectively, may take years to complete
Randomised clinical trial	Most convincing design Only design which controls for unknown or unmeasurable confounders	Most expensive Artificial Logistically most difficult Ethical objections

Hypothesis development is a critical step in the scientific process. Hypotheses are used to make predictions, which are then tested by further research. If the test results are consistent with the hypothesis, the likelihood of the hypothesis being true is strengthened. If the results are not consistent with the hypothesis, it needs modification.

In terms of a hierarchy of study designs for evaluating the effectiveness and cost-effectiveness in clinical practice of drug interventions, the pragmatic RCT conducted in "real world" conditions is typically viewed as the gold standard**.** This is because they have high external validity (i.e. establishing whether the costs and effects of an intervention are representative in clinical practice), which is a key element for maximising the value of economic evaluations for decision making and, if well designed, retain key aspects of internal validity. RCTs (double blind placebo controlled) in which a defined treatment protocol is followed represents the gold standard for measuring efficacy, with phase III clinical trials being strictly controlled through tightly defined criteria designed to meet the requirements of drug licensing authorities such as the Food and Drug Administration in the USA and the European Medicines Agency in Europe. These RCTs are in general second best options in terms of their design for the economic evaluation of pharmaceuticals, as despite high internal validity (important for minimising bias in any evaluation), there are limits on external validity (e.g. compliance is often much higher than in real clinical practice) .

Observational designs represent another second level source of primary data for economic evaluations. They should have good external validity if based on treatment and care received in practice, but have potentially severe limitations in internal validity.

If such designs were being used to compare the cost-effectiveness of two or more interventions, there could be a high risk of bias, random error and confounding factors when comparing groups receiving the different drug or other interventions being evaluated.

Bias, also known as differential error, is a dangerous source of error. Bias usually produces deviations or distortions that are usually systematic and in one direction. It may occur when there is an unequal allocation of subjects, unequal detection of the outcome, or unequal loss to follow-up of subjects in the various patient groups being studied. Bias becomes a problem when it weakens a true association, produces a spurious association, or distorts the apparent direction of association between variables.

Random error, also known as differential error, produces results that are too high or too low in approximately equal amounts, owing to random factors. Even though, it is a serious problem, random error is ordinarily less serious than bias because it is less likely to distort (i.e reverse the direction of) the findings. It does, however, decrease the probability of finding a real association by reducing the statistical power of a study.

Confounding from the latin 'to pour together' is the confusion of two supposedly causal variables, so that part or all of the purported effect of one variable is actually due to the other.

An observational study design could potentially provide all of the core resource use consequence, clinical event and outcomes data for a full pharmacoeconomic evaluation. There are a number of issues concerning whether this represents a feasible prospect, relative to the use of RCTs (pragmatic or explanatory):

Validity: Pragmatic RCTs have high external validity, RCTs have highest internal validity. Observational datasets may be considered to perform less well on both counts. Therefore, in terms of overall validity pragmatic RCTs score higher. However, if the choice was between using an RCT or an observational design, the latter may well perform better on external validity, and if this is considered more important, it could outweigh the cost associated with loss of internal validity. However, it is not sufficient to ignore the problems of bias and confounding factors – so that if an observational design is used fully for an economic evaluation, steps should be taken to ensure that there is a non-spurious relationship between drug intervention and cost/effect outcomes.

Comparators: In economic evaluations the drug(s) being evaluated should be compared to an alternative drug or treatment representing current best practice. Pragmatic RCTs are specially designed to compare different interventions, whilst controlling to some extent for bias and confounding factors. A limitation of non-pragmatic RCTs is that the comparator is frequently only placebo, which is not ideal for economic evaluations designed to measure drug performance against best alternative treatments. Prospective observational study designs may be constructed to address a specific economic evaluation research question with groups of patients identified to receive different drug treatments. However, retrospective or semi-

retrospective databases are not typically set up for specific evaluations, but designed to collect as much relevant data as possible on resource use, clinical events and outcomes (depending on their scope and detail) for a disease or condition in one of more settings (e.g. hospital, primary care). Whilst optimal for reflecting resource use and outcomes in actual practice, there may be difficulties in identifying comparable sub-groups of patients taking the drug and comparator treatments, and independently assessing costs and outcomes for each. Comparison bias is the main limitation of observational studies because of the lack of randomisation (Byar, 1991).

Complexity and design: Phase III RCTs are not typically designed for economic evaluations, hence there are usually limitations on the amount of resource use and QoL/health status outcomes that can be collected. In addition, RCTs may not be ideally appropriate for addressing the economic research question, if for instance the follow-up period chosen to measure efficacy is insufficient to assess the full effect of the drug on costs and QoL outcomes. For example, the efficacy of a stroke drug such as tPA may be measured as 3 month change in Modified Rankin score (a measure of post stroke disability), whereas for cost-effectiveness at least a one year follow-up would be required to measure impact on such outcomes as nursing care costs and QoL impact**.** Therefore longitudinal/cohort databases allow researchers to trace care over time. Observational designs in general offer an alternative method of collecting sufficient data for a cost-effectiveness analysis, and if designed prospectively to address a specific research question for a societal (or more limited) perspective, ensure the appropriate resource use, clinical event and health outcomes data are collected. Also such designs allow little or no intervention by study coordinators in the process or delivery of care to patients (Strom et al. 1990), i.e. it is non-intrusive. Retrospective datasets may have limitations in terms of the detail contained on each of these core variables to address all the CEA/CUA research questions from different perspectives. For example, indirect cost data may not be available or utility/QoL outcomes not recorded and so CUAs of a drug intervention could not be based fully on such a dataset.

Generalisability: A full economic evaluation based on a phase III multinational RCT may have poor generalisability to any particular country setting for reimbursement/prescribing decisions. Numbers of patients from any one country in the RCT may be too low to enable reliable economic evaluations in that country setting (Willke et al., 1998). An RCT (either pragmatic or explanatory) carried out in one country or in only a few countries may provide more representative data for a specific country context. Study specific observational designs would suffer from the same problems unless designed to collect sufficient and standardised patient data from a number of sites across selected countries. Prospective datasets, such as those available for the UK (see section 7) tend to have a large number of patients collected from a number of sites across the country and so contain data that is potentially representative and generalisable within the country.

Convenience and cost: Retrospective observational datasets score highly on speed and convenience. They are usually available, at a cost, for immediate access by researchers. Full economic evaluations can therefore be potentially carried out quickly, compared to the time it takes to set up a pragmatic RCT, or prospective observational

study. Explanatory RCTs are not usually set up for economic evaluations, which tend to "piggyback" onto these studies. Therefore, it may be convenient to include economic and outcomes data within a phase III clinical trial. However, speed is still an issue here in that the results will not be available until completion of the trial, which may take one or more years depending on the disease, aim of the study, patient recruitment rate and follow-up period. Commercially available retrospective datasets are likely to have relative advantages in cost for the study sponsor and are continually updated. In general, policy makers need to make the best decisions they can today, recognising that data are not (nor ever will be) perfect. Furthermore the dynamic evolution of many healthcare technologies is such that the contemporary policy question may have changed in the time it takes to gather primary data. In the case of defibrillators, it is only recently that randomised studies have been undertaken (Moss et al., 1997). Data from one-off studies become outdated quickly, and are costly to update. Observational datasets with ongoing data collection and monitoring can be readily used to update the results of cost-effectiveness analyses. Non-commercial ventures (at least originally) to set up semi-retrospective datasets such as the National Prospective Monitoring System (NPMS) for HIV/AIDS (a bottom-up resource use and outcomes dataset collecting standardised data across 10 hospital sites in England), or the Sheffield ICU database (a single hospital patient dataset on costs of ICUs) are costly to run, and face particular funding difficulties often looking for industry sponsorship in order to continue data collection. Further there is often a lag-time before data availability. Data capture may have taken place for example, up to 2 years before analyis.

The above issues relate to the relative advantages and disadvantages of observational designs for full economic evaluations. The most important problem is that of bias and confounding factors (Sheldon 1996). One approach to minimise this problem and to still have all the benefits of external validity, speed, low cost and convenience) is to use observational study designs to provide partial data for an economic evaluation. Prospective and retrospective studies could supply comprehensive data on resource use and (possibly) outcomes reflective of real clinical practice, with clinical efficacy and some outcomes data derived from phase III clinical trials [*what the review showed here*], meta-analyses or systematic reviews. Retrospective observational datasets that are available tend to adequately cover resource use data, and are less useful health outcomes data such as QoL outcomes – hence these could be used for supplying resource use information for cost-effectiveness models of drug therapies.

Alternatively, multivariate modelling methods could be used to better adjust for confounding factors and bias in observational datasets used for full economic evaluations. The potential or using statistical methods are dealt with in section 6.

Resource requirements: Observational datasets have benefited from recent advances in hardware and software. However retrieval of data requires knowledge of programming, epidemiological expertise and of the design and structure of the database. Several datasets require a working knowledge of SAS or SQL.

In brief, some of the general features an ideal database would have are as follows:

- Data on demographics

- Date on morbidity, mortality, prescribing and treatment
- Population-based
- Large size
- Geographical representative
- User friendly
- Timely access
- State of the art analytical capability
- Tool to support innovative research
- High quality and validated

6. Methods to Analyse Longitudinal Observational Data

Statistical techniques, which come under the broad heading of multivariate analysis, have been developed to analyse data gathered through the observation of many variables. These techniques help to deal with situations where it is impossible to control the generation of appropriate data through experimental design.

Multivariate analysis allows you to elicit information from observational data sets that contain simultaneous measurements on many variables. In this case the variables of interest may be health care resources and health outcomes for large groups of patients.

Below are some suggested objectives of multivariate statistical investigations of observational data sets:

- *Data reduction or structural simplification to make interpretation easier*. As several variables may relate to a patient's response to a particular treatment the data can be simplified to construct a simple measure of patient response.

- *Generate and apply rules for classifying study subjects into groups.* Observational data sets could be used in economic evaluations to classify patients into treatment groups, e.g. drug A or drug B, and health outcome groups, e.g. treatment success or treatment failure.

- *Investigate the dependence among variables.* Health outcomes will be dependent on a number of observed and unobserved variables.

- *Compensate for bias due to the absence of experimental design.* Economic evaluations alongside randomised controlled trials are regarded as the gold standard as the primary objective of the design of these experiments is to eliminate statistical bias. Statistical techniques can be used to control for some of the inherent bias of data gathered in observational data sets.

In order to achieve these and other objectives the following approaches could be adopted:

- Generalised linear modelling

- Principal component analysis
- Factor analysis
- Discriminant analysis
- Cluster analysis

Although it has been suggested that use of these techniques can be controversial (Hlatky, 1988), applications are beginning to appear in the literature (Croghan et al. 1997). Advances in computer technology and the development of sophisticated statistical software packages have the made the implementation of these and other multivariate statistical techniques for data analysis more straightforward. In addition, econometric tools are being developed to analyse situations in which no control group is available for comparisons (Mennemeyer, 1997) and multiple imputation methods for incomplete data (Statistical Solutions, 1997; Rubin, 1978).

Cross-design synthesis is an innovative technique put forward by the US General Accounting Office (1992) that combines results from diverse, but complementary studies that evaluated a given treatment's effect. Specifically, it can be used to assess, adjust and combine treatment effects obtained across RCTs and observational database analyses. Cross-design synthesis has some limitations. For example, the method may involve statistically complex adjustments and overlook the fact that inappropriate data are being pooled. In a related approach, Eddy et al. (1992) have also presented the confidence profile method as a set of techniques for adjusting and combining data from mutiple sources.

Table 4: Cross-Design Synthesis: Complementary Strengths and Weaknesses of Study Designs for Assessing Effectiveness

Study Design	Primary Strength	Primary Weakness
Randomised Studies	Controlled comparison Internal validity	Potential lack of generalisability External validity at risk
Database Analyses	Coverage of medical practice (full payment population, full range of treatment implementations); External validity	Uncontrolled comparison Internal validity at risk

Source: US General Accounting Office, 1992

7. Developments in Observational Databases

The majority of observational databases have not been developed specifically for economic evaluations. In section 3 we outlined three types of observational dataset. An important development for the provision of data for pharmacoeconomic evaluations, has been in the construction of prospective datasets with wide patient coverage for a country (therefore, increasing data representativeness) and which are available, either commercially or publically, for use by academic or industry health economists and outcomes researchers. In this section some brief details of general primary/secondary databases (7.1), and disease/setting specific databases (7.2) for the UK that fit into this category are presented.

7.1 General Databases

General databases may often contain limited information on clinical/health outcomes, and no data on QoL or utility outcomes (for CUAs), but have the advantages of a large numbers of patients and a high external validity. Four main general automated databases for the UK were reviewed by the Centre for Medicine Research (CMR) International (Lis et al, 1997): GPRD, Memo, DIN-LINK, Mediplus. The first two of these were set up originally to provide supplementary data for the evaluation of drug safety, whilst the latter two were for marketing purposes. Memo is for Scotland, whilst the other three cover the whole of the UK with representation from 3-11% of the patient population. Although not originally designed for Outcomes Research, they all contain useful core data for pharmacoeconomic evaluations. Table 5 contains summary details of all four.

7.2 Specific databases

Specific prospective databases, such as those outlined in this section and in table …. contain predominantly data on resource use, with some data on clinical events, and health outcomes. They may contain data on QoL/utility outcomes. The resource use and outcomes data should have enough external validity in order to be used as an effectiveness measure or to be extrapolated to cost-effectiveness in pharmacoeconomic studies. The numbers of patients in a specific database is generally lower than in the general databases outlined in section 7.1 so the measurement of statistical significance may be more difficult. Two specific databases are included in table 5. First, the Medical Economics and Research Centre (MERCS) database on ICU resource use and costs, based currently on a single ICU site in Sheffield but with plans and aspirations to develop into a multi-centre patient specific cost (and outcome?) database (and ultimately multi-national). Second, the National Prospective Monitoring System for HIV/AIDS has been running since 1994, and has grown from a single London hospital clinic to collect standardised hospital resource use and outcome data from about 13 HIV Units in England. Possible future developments with this database is to expand to cover community care and develop similar datasets in other countries.

Both databases have good internal validity check procedures (e.g. for dealing with missing data), are very detailed in their resource use coverage and have been used in published cost analyses (e.g. for Sepsis costs in ICU – see Edbrooke et al, 1999, and Beck et al 1998 for the use of the NPMS to estimate the costs of HIV hospital care for England), but not yet partially or fully in CEAs or CUAs of pharmaceuticals (or any other intervention).

Table 5: Examples of general and specific databases containing core data of potential use for pharmacoeconomic studies

	General databases*				Specific databases	
Name of database	**DIN-LINK**	**GPRD**	**Mediplus**	**Memo**	**NPMS for HIV/AIDS**	**MERCS ICU costs database**
Number of patients	800,000-3 million	3.5-6 million	1.6 million	400,000+	50% of diagnosed HIV population in 1997	Over 1000
Geographical Coverage	UK	UK	UK	Scotland	England	Sheffield and Trent Region
Number of General Practices or hospitals	100-210 GPs	500 GPs	140 GPs	All Tayside area GPs & Hospitals	13 HIV Clinics (GU clinics)	One large ICU
First data recorded	1987	1987	1991	1980 (hospital), 1989 GP	1994	1995
Scope of main dataset	Patient demographicss, prescriptions, diagnosis, symptoms, tests, 2ndary referrals, main outcome of referral			Patient demographics, prescriber details, prescriptions, detailed secondary care	Patient demographics, risk factors, inpatient, outpatient, day patient use (and clinical outcomes), procedures, tests , drugs, diagnosis,	Patient demographics, severity, diagnosis, inpatient stay and consultation cost, resource use/cost for: admin, drugs, crystalloids, tests, procedures, treatment, ward rounds, nursing
Frequency of data collection	Daily	6 weekly	Daily	Continuous or as required	3 monthly	Continuous

*Source: Lis et al (1997)

8. Discussion and Conclusions

Regardless of the system of health care, there is an obvious and universal need for reliable data on the effectiveness and costs of clinical interventions. RCTs have been long regarded as the only method for evaluating drug effects. With the dramatic growth in computing power, observational methods have increased progressively in importance and the use of experiments and observations as complementary tools is gradually gaining the acceptance of researchers. Along with this growth has been the increasing concern over financing health care across large populations, which has increased the demand for information on how health care technologies perform in large populations of heterogenous patients.

The choice of experimental or observational data inevitably represents a series of compromises, but the challenge is to develop a new strategy that will be capable of generating results that have an acceptable balance between internal and external validity.
Observational datasets have implications for outcomes research conducted in pharmaceutical companies. In an increasingly austere regulatory environment where economic implications of compounds are being assessed, outcomes research scientists will have to look for ways to satisfy the needs and requirements of key decision-makers. To be able to demonstrate the clinical and economic benefits necessitates decisions on the choice of experimental or observational data within the constraints of resources and time have to be made. Inevitably this choice represents a series of compromises and the challenge is to develop a strategy that will be capable of generating results that have an acceptable balance between internal and external validity, between specific accuracy or broader policy relevance and between immediate solutions and long-term study.

Currently, the comparison of randomised and non-randomised designs has been inconclusive. In the US, Patient Outcomes Research Teams have attempted to compare the results of observational studies with those of clinical trials. The results have been inconclusive; 'Sometimes non-randomised studies will tell you the right answer, sometimes the wrong answer, and there is no way to tell the difference without an RCT to determine the true answer' (US Congress Office of Technology, 1994). Despite this controversy, it is less deniable that the observational data are useful in filling the void regarding detailed economic information and generalisable conclusions, and the larger prospective datasets now available have been a relatively underutilised resource available to Outcomes Researchers (Lis et al, 1997). The Lis et al report presented results from interviews with Outcomes Researchers working in drug companies in the UK who identified the value of observational datasets for the production of burden/cost of illness studies. They also have clear potential value in the provision of partial data for a pharmacoeconomic evaluation, in particular the provision of resource use data to complement clinical event/outcome and health/QoL outcomes from RCTs, special research question specific observational studies, or other data sources. However, there is little evidence of studies using any of the large longitudinal prospective datasets such as those in table 5 (e.g. see the review in section 4). For example, the Sheffield ICU database could be useful for costing treatment pathways and clinical events for a pharmacoeconomic study alongside a phase III clinical trial, but has not as yet been used for such purposes. This may be because Outcomes Researchers in drug companies under-estimate the complexities of resource use measurement in ICU

patients, and so do not see the need for accurate ICU costs data of the sort that the Sheffield database could provide.

There was no evidence from the literature review in section 4 of the use of prospective observational datasets to provide the full set of core data items for pharmacoeconomic studies. Despite the potential for this, there are a number of barriers that would need to be overcome if more drug company sponsored studies were to be undertaken using such data.

Firstly, drug companies would need to recognise the value of such datasets for pharmacoeconomic evaluations. For this to occur they would need better marketing of the dastasets (which is improving as the data providers become more streetwise) and for the data to be more comprehensive in certain areas (e.g. outcomes data, scope of resource use data) and to further increase representativeness by supplying standaredised data from more centres (a problem particularly faced by specific databases which in order to get centres to subscribe to providing data rely on a lot of goodwill and encouragement that that data will be useful for local audit and budget claims purposes). It is a bit of a chicken and egg situation. To provide the extra data requires resources, which requires the databases to be sold to drug companies – the NPMS and the ICU databases are currently examining arrangements as to how to increase the level of industry sponsorship in order to further develop the databases.

Secondly, the best multi-variate techniques for reducing bias and confounding factor problems will need to be further explored.

Thirdly, none of the main guidelines for cost-effectiveness analysis of new drugs for reimbursement purposes that exist for different countries explicitly state whether observational study designs are an acceptable source for pharmacoeconomic studies that are submitted to support a reimbursement or pricing claim. The Ontario guidelines recommend data based on meta-analyses of RCTs. The Australian guidelines strongly recommend the use of results from randomised clinical trials, supplemented by additional information. The Canadian guidelines do not specifically address this issue, but appear to favour data from clinical trials. The British guidelines (ABPI'94), US Hillman, US PhRMA , Belgian, and German guidelines are least prescriptive in this regard and mention results from various types of studies. The impending guidelines from NICE might give a clearer idea of the acceptable study design expected of pharmacoeconomic studies for new drugs in the UK. However all sets of guidelines recognise that RCTs measure the impact of an interventions under ideal conditions (efficacy) rather than under conditions of normal practice (effectiveness).

There is little literature on the use of observational data in economic evaluation, and more debate on this subject is needed. As we gain a better understanding of the tradeoffs between internal and external validity, it is predicted that more pharmacoeconomic analyses will increasingly be based on observational data or a combination of observational and RCT data.

This is very much work in progress and designed to kick start a debate. The authors would welcome you to share your thoughts on the usefulness of observational datasets for economic evaluation.

Address for Correspondence: Keith Tolley, Outcomes Research, Pfizer Central Research, Ramsgate Road, Sandwich, Kent CT15 9NJ. E-mail: keith_tolley@sandwich.pfizer.com

References

Adams, M.E., McCall, N.T., Gray, D.T. et al. Economic analysis in randomized controlled trials. *Medical Care* 1992;30:231-243

Beck, E, Tolley K, Power A et al for the NPMS Steering Group. Use and cost of HIV service provision in England 1996. *PharmacoEconomi*cs 1998, 14:639-52.

Byar, D.P. Problems with using observational databases to compare treatments. *Statistics in Medicine* 1991; 10:663-666

Commonwealth Department of Human Services and Health. *Guidelines for the pharmaceutical industry on preparation of submissions to the pharmaceuticals benefits advisory committee: including major submissions involving economic analysis.* Canberra: Australian Government Publishing Service, November 1995

Croghan, T.W., Lair, T.J., Engelhart, L. et al. Effect of antidepressant therapy in health care utilization and costs in primary care. *Psychiatric Services* 1997.

Drummond, M.F., Davies, L.M. Economic analysis alongside clinical trials: revisiting the methodological issues. *International Journal of Technology Assessment in Health Care* 1991;7(4):561-573

Drummond, M.F. Experimental versus observational data in the economic evaluation of pharmaceuticals. *Medical Decision Making* 1998;18(2): S12-S18

Edbrooke D, Hibbert C, Ridley S, Long T, Dickie H. The development of a method for comparative costing of individual intensive care units. *Anaesthesia* 1999;54(2):110-120.

Edbrooke D, Hibbert C, Kingsley J, Smith S, Bright N, Quinn J. The patient-related costs of care for sepsis patients in a United Kingdom adult general intensive care unit. *Critical Care Medicine* 1999; 27 (9):1760-67.

Elixhauser, A., Halpern, M., Schmier, J., Luce, B.R. Health care CBA and CEA from 1991 to 1996: An updated bibliography. *Medical Care* 1998; 36(5):MS1-MS9

Gold, M.R., Siegel, J.E., Russell, L.B. et al. editors. *Cost-effectiveness in health and medicine.* New York: Oxford University Press, 1996

Hillman, A.L., Eisenberg, J.M., Pauly, M.V. et al. Avoiding bias in the conduct and

reporting of cost-effectiveness research sponsored by pharmaceutical companies. *New England Journal of Medicine* 1991; 324 (19):1362-1365

Hlatky, M.A., Califf, R.M., Harrell, F.E et al. Comparison of predictions based on observational data with the results of randomised controlled clinical trials of coronary artery bypass surgery. Journal of American College of Cardiology 1988;11:237-245

Jonsson, B. Economic evaluation of health care technologies. A review paper. *Acta endocrinologica* 1993;128(2):50-54

Kalish, S.C., Gurwitz, J.H., Krumholz, H.M. et al. Cost-effectiveness of model of thrombolytic therapy for acute myocardial infarction. *Journal of General Internal Medicine* 1995;10:321-330

Lis Y, Carter M, Lumley C, Walker S. Can the UK automated databases provide relevant information for health economics research. CMR International Report, 1997.

Mennemeyer, S.T., Can econometrics rescue epidemiology. *Annals of Epidemiology* 1997; 7(4):249-250

Moss, A.J., Hall, W.J., Cannon, D.S. et al. MAIDIT Trial. *New England Journal of Medicine* 1997; 335:1933-1940

NICE. *Faster access to modern treatment.* Discussion paper, HMSO, London 1999a

Nixon, J., Pang, F. Economic evaluations in Japan: A review of published studies, methodological issues and practice, *forthcoming paper* 2000.

O'Brien, B. Economic evaluation of pharmaceuticals. Frankenstein's monster or vampire of trials. *Medical Care* 1996; 34(12):DS99-DS108, Supplement

Rittenhouse. B.E. The standards debate. What have we learned ? *Medical Care* 1996;34(12):DS5-DS10, Supplement

Rubin, D.B., Bayesian inference for causal effects: the role of randomization, *Annals of Statistics* 1978; 6(1):34-58

Sackett, D.L., Haynes, R.B, Tugwell, P. *Clinical epidemiology: A basic science for clinical medicine.* Little, Brown and Co., Boston 1985

Schwartz, D., Lellouch, J. Explanatory and pragmatic attitudes in therapeutic trials. *Journal of Chronic Diseases* 1967; 20:63-64

Sheldon, T.A. The problems of using modelling in the economic evaluation of health care. *Health Economics* 1996;5;1-11

Sloan, F. ed. *Valuing health care.* Cambridge: Cambridge University Press 1995

Statistical Solutions. *Solas for missing data analysis: the solution for missing values in your data. User reference manual*. Cork, Ireland, Statistical Solutions, 1997.

Strom, B.L., Carson, J.L. Use of automated databases for pharmacoepidemiology research. *Epidemiological Review* 1990;12:87-107

United States General Accounting Office. *Cross-design synthesis: A new strategy for medical effectiveness research.* Washington DC: General Accounting Office 1992; GAO/DEMD-92-18

US Congress Office of Technology Assessment. *Identifying health technologies that work: searching for evidence (OTA-H-608)*. Washington DC: US Government Printing Office, 1991

Willke, R.J., Glick, H.A., Polsky, D. , Schulman, K. Estimating country-specific cost-effectiveness from multinational clinical trials. *Health Economics* 1998; 7:481-493