Economic evaluation of pharmaceuticals: what are reasonable standards for clinical evidence—the Australian experience

Suzanne Hill,[1] David Henry,[1] Brita Pekarsky[2] & Andrew Mitchell[3]

[1] *Pharmaceutical Policy Analysis Group, Discipline of Clinical Pharmacology, Faculty of Medicine and Health Sciences, The University of Newcastle, New South Wales,* [2] *Department of General Practice, School of Medicine, The Flinders University of South Australia, South Australia, and,* [3] *Pharmaceutical Evaluation Section, Department of Health and Family Services, Canberra, Australia*

Keywords: pharmacoeconomics, Australian guidelines, cost-effectiveness analysis.

Introduction

In many countries increasing expenditure on pharmaceutical products is coming under intense scrutiny by Governments. In 1993, the Australian government introduced a requirement that, in order to have drugs listed on the national pharmaceutical products reimbursement scheme (the Pharmaceutical Benefits Scheme), pharmaceutical companies must submit formal economic evaluations of their products, providing evidence of their comparative effectiveness, cost effectiveness, estimated use and total financial cost to the federal government. Australia was the first country to introduce an explicit requirement for economic data in relation to subsidization of pharmaceuticals, beyond the normal regulatory requirements of quality, efficacy and safety. This has been described by Drummond [1] as the first example of the 'fourth hurdle' and it has provoked considerable debate.

One of the key criticisms of the Australian approach to measurement of cost-effectiveness of new pharmaceutical products is that clinical issues are seen by some economists to dominate the evaluation, perhaps at the cost of broader economic or health system considerations [2]. A particular concern to critics has been the emphasis given in the Australian guidelines [3] to randomized trials as the basis of judgements regarding the relative performance of new and established drugs [2, 4]. This criticism raises a fundamental issue about the requirements for evidence in assessing the cost-effectiveness of pharmaceutical products: should the evaluation be based principally on the results of randomized trials, or can we rely on the results of observational studies and data derived from large administrative data-bases?

In this article we argue that real world decision-making does not involve 'system-wide' judgements. Instead decisions regarding new drugs are made on the margin by comparing them with existing remedies for the same indications. Because of a lack of final outcome data when these decisions are made, comparisons about value for money across and between disease states are difficult, but not impossible. The first logical step is to establish the true clinical performance of the new drug against an agreed comparator, preferably using data from randomized comparative clinical trials. After the clinical role has been established the economic performance can be assessed, sometimes using data from the clinical trials, and often from some straightforward economic modeling using the clinical trial data as the basis of determining the differences in clinical outcomes.

Correspondence: Dr. S. Hill, Discipline of Clinical Pharmacology, Faculty of Medicine and Health Sciences, The University of Newcastle, Callaghan, New South Wales 2308, Australia.

In this paper we consider some of the issues regarding the standards of evidence that are appropriate for pharmacoeconomics, and the thinking behind the approach that has developed in Australia.

Background

The Australian health system is a combination of a publicly funded system and a private, 'user pays' system. There is a national medical insurance scheme (Medicare) and within it, a national reimbursement scheme for pharmaceuticals—the Pharmaceutical Benefits Scheme (PBS). The PBS was introduced in 1953 with the aim of providing access to essential drugs for all Australian residents, and has remained relatively constant through the life of several governments. It is a 'positive' formulary, with decisions made to list drugs, rather than to black list them (in contrast with the UK). Effectively, if a prescription drug for non-hospital use is not on the PBS, it does not have a market in Australia.

Before a drug can be considered for inclusion on the PBS, it must be licensed/registered in Australia. A regulatory authority, the Therapeutic Goods Administration (TGA) is responsible for registering drugs. Companies submit registration applications to the TGA, the applications are evaluated and then the Australian Drug Evaluation Committee (an independent advisory body) provides clinical advice about whether the application should be approved. The decision to approve or reject the application for registration is then made by the TGA.

The data required to support an application for registration are similar to those required in Europe. The TGA must be satisfied that the product is of adequate quality, safety and efficacy for use in Australian medical practice, and therefore, the standard type of pharmaceutical data, toxicology studies and clinical trials are required. The number and design of clinical trials is not specified, although most companies submit randomized placebo-controlled trials, as required by the American Food and Drug Administration. The TGA does not consider the cost of the product. In general, regulatory decisions tend to be congruent with those of major regulatory agencies, although there are some consistent differences in thinking. For example, very few fixed combination products are approved for marketing in Australia.

While the registration package is being evaluated, the sponsor company may lodge an application for it to be listed on the PBS. This application is made to the Department of Health, and contains data that are evaluated by Departmental staff as well as two advisory committees. The first review is carried out by the Economics Sub-committee (ESC), which

considers the technical aspects of the submission, particularly the economic analysis. The second review is by the Pharmaceutical Benefits Advisory Committee (PBAC), which considers advice from the ESC and makes a judgement on whether the drug seems to represent 'value for money'. The legislation requires that the PBAC consider the effectiveness, cost-effectiveness and clinical place of any new product, comparing it with other interventions for the condition, including non-pharmacological treatments.

Once a drug is recommended for listing, a final price is negotiated between the sponsor company and the Pharmaceutical Benefits Pricing Authority, who base their negotiations on the advice of the PBAC. The drug then appears in the Schedule of Pharmaceutical Benefits, can be prescribed by registered medical practitioners and the Commonwealth Government will subsidize the use of the drug. The Schedule includes different levels of availability that are determined by the recommendation of the PBAC. If a drug is placed in the most restricted category, a telephoned application for 'authority' to prescribe the drug for an individual patient is required. Typically the doctor will be required to provide details of the diagnosis, how it was made, and (if relevant) the date of any diagnostic test. The authority system is administered by the Health Insurance Commission, responsible for payments to doctors, and is often used to direct the subsidy to the patient groups for whom the drug is likely to be the most cost-effective alternative.

The data that a company needs to submit to support a listing application are set out in the Guidelines for the Pharmaceutical Industry on Preparation of Submissions to the PBAC [3]. These were first released in draft form in 1990, revised and formalized in 1992 after input from industry and then revised again in 1995. The sponsor company is required to relate the clinical outcomes of treatment with the new drugs to the costs associated with its use and to compare these with the estimated costs and outcomes of treatment with the drug that it is most likely to replace. Included in the analysis are the acquisition costs of the drug, additional costs of administration and monitoring, and the costs of hospitalization and additional clinical care, and where available, estimates of the impact of the change in quality of life that may result from an effective new drug. The latter are seldom available at the time the PBAC first considers a new product, and as a result, it is often necessary to base judgements on the effects of a drug on surrogate outcomes. Changes in productivity due to reduced sick leave are accepted, but only if supported by reasonable evidence. There are a number of controversial issues regarding the correct way of incorporating costs in a cost-effectiveness analysis. However this article is concerned mainly with the analysis of clinical outcomes.

Cost minimisation or cost-effectiveness?

A company can adopt one of two basic approaches in making a case for its new product. These are either to claim *equivalence* of the new product compared with one that is already available, or to claim clinical *superiority*. More recently, there have been examples where the sponsor accepted that a drug was clinically inferior, but was cheaper. As discussed by Laupacis *et al*. these examples are uncommon, and difficult to assess, but their inclusion may be 'rational' from an economic standpoint so long as the savings are great enough to justify the loss of benefit [5].

If a company wishes to claim that its product is equivalent, and therefore adopt a cost-minimisation approach in the economic analysis, the first question that needs to be answered is whether the products are truly equivalent. Cost minimisation analysis requires that the efficacy of the two products be the same. The net costs associated with the new drug should be the same, or less than those of the comparator. As administrative costs, and costs of monitoring, tend to be similar for equivalent drugs, this usually means that the new drug will be awarded the same price as the comparator.

To establish clinical equivalence, the magnitude of the effect of each drug must be determined. Because any differences in efficacy are likely to be fairly small this can only be done on the basis of randomized controlled trials. Until relatively recently, active comparator trials have been the exception rather than the rule, with most trials of new drugs being placebo-controlled. Where active comparator trials have been conducted, most have been designed to show differences between treatment groups rather than equivalence; the sample size required to establish equivalence with an acceptable degree of precision is usually larger than that required to show clinically significant differences [6].

It is possible for drugs to appear to be equivalent in conventional statistical terms i.e. the P value for a comparison of means is greater than 0.05, but at the same time, for one of them to be clinically inferior to the other, simply because the sample size of the trial is too small to detect the difference. As with bioequivalence studies the preferred approach is to carry out a confidence interval analysis. But what degree of precision is desirable? This will vary with the clinical setting. Establishing true equivalence may seem to be most important for drugs that are used to treat uncommon life-threatening diseases. On the other hand small differences in the efficacy of drugs used to treat hypertension, lipid disorders or depression could translate into thousands of under treated cases and the actual burden of preventable disease resulting from subsidisation of inferior drugs could be even greater. It is regrettable that the comparative trials of these agents are seldom large enough to detect clinically significant differences.

If a company claims *superiority* for their new product, the clinical trials and economic analysis need to demonstrate the extent of superiority of the new treatment over the old, and provide a valuation of that difference. The magnitude of any difference in treatment effect tends to be a major determinant of the outcome of any economic analysis. Cost-effectiveness analysis, cost-utility analysis, or cost-benefit analysis can be used as a basis for this assessment. The PBAC will then decide whether it agrees with the company's estimate of the advantages of the new product, and the clinical economic and social valuation placed on them. If it is the first product for a new indication, the price recommended for the new product will be likely to become the benchmark for subsequent products for this indication. As with cost-minimisation, establishing the relative performance of new products can be difficult in the absence of adequately powered, active comparator trials.

Reliance on randomized trials

The PBAC has developed a hierarchy of preferred evidence that can be used to support claims of equivalence or superiority. In the experience of the committees, both claims are best supported with an economic analysis that is based on the results of randomized controlled trials that have the comparator as the control. In cases where no 'head-to-head' trials are available, the Guidelines have encouraged the use of 'common comparator' trials, where both the new drug and the comparator are tested against a common reference therapy, often placebo. Assessing equivalence or superiority on the basis of this type of data is very difficult. The data in Table 1 represent a hypothetical comparison between two active drugs where only placebo-controlled trials are available. The drugs are being used to prevent an unspecified outcome. The difference in the risk of the outcome in treated *versus* placebo-treated subjects in Trial 1 is 0.10, while in Trial 2 the difference is 0.05. Does this mean that Drug A is more effective in reducing the frequency of the outcome than Drug B? Can we infer that the difference in outcomes with Drug A compared with Drug B is 0.05 (0.10–0.05)? Should Drug A be awarded a higher price on the grounds of its apparently greater efficacy?

Note that in both trials the effect of treatment was to *halve* the rate of the outcome. The Relative Risk (RR) in both cases is 0.50. In such circumstances the RR is a more reliable estimate of treatment effect than the risk difference, as it not influenced by the background (placebo) rates, which differ between the two studies. In this case the safest assumption is that there is no convincing evidence of a difference in efficacy between Drugs A and B. This situation often has to be faced when making decisions regarding newly marketed compounds. Because of the problems of interpreting such non-comparative data the committee has developed a 'pragmatic' policy of adjudicating claims regarding equi-effectiveness, but has found it difficult to accept claims of superiority in the absence of comparative trials. A recent example that has been highlighted in the published literature is the plethora of new anticonvulsant drugs [7]. These have all been trialled as 'add on' treatments in patients with severe epilepsy uncontrolled with conventional treatments and most of the trials have employed placebo controls. As Marson points out, the apparent variability in treatment effects when these agents have been compared with controls is not a satisfactory basis for drawing conclusions regarding the *relative* performance of the drugs—such inferences have to be based on comparative trials.

Table 1 Comparison of two treatments through a common comparator.

Consider the results of the following trials. Two therapies intended to prevent an outcome (Treatment A and Treatment B) are being compared, but there are no direct comparative 'head to head' trials. However there are two placebo-controlled trials. The following results were obtained:

	Trial 1		*Trial 2*	
	Treatment A	*Placebo*	*Placebo*	*Treatment B*
Outcome	10	20	10	5
No outcome	90	80	90	95
Total	100	100	100	100

Note: We are ignoring for convenience issues of statistical precision, so confidence intervals have not been calculated.

Another major criticism of the use of randomized trials is that high 'external validity' is necessary when performing pharmaco-economic analyses. It is argued that this can only be achieved at the cost of 'internal validity'. In other words, it is assumed that attempts to control biases in a trial reduce the generalisability of the results, and that in pharmaco-economics it is necessary for a trial to have high external validity in order to apply the results to a 'representative' population of potential users. There are two issues to consider here. The first concerns the estimation of the size of the *clinical effect* when the drug is used outside the conditions of the trials. The second relates to applicability of the data on *costs* that are collected as part of the trial.

There are some common misconceptions regarding the use of clinical trial results to estimate the benefit of a drug when it is used in real life. The traditional approach to application of clinical trial results has been to compare the characteristics of the trial participants with those of the reference population of interest, and to pay close attention to patient selection factors, inclusion and exclusion criteria [8–10]. It is widely assumed that these factors determine the way in which the results should be generalised. It is important to remember that the results of clinical trials reflect the *average* effect of a given intervention used in the study population of subjects. Even if the average effect is positive there will be subjects in the study population who experienced a greater than average benefit, and others for whom the treatment did more harm than good [11]. In many situations the *relative* treatment effect is fairly constant across different levels of baseline risk. A review by Schmid *et al*. has indicated that constancy of relative treatment effects seems to hold true across a wide range of important therapeutic questions [12]. Examples are: reduction of blood pressure and major cardiovascular events and the effects of warfarin in non-rheumatic atrial fibrillation on embolic stroke [13, 14]. The importance of these observations is that the magnitude of the treatment effect can be estimated reasonably from a knowledge of an individual's baseline risk [15] and the relationship that holds between the baseline risk and the magnitude of the treatment effect. In other words, the accurate generalisation of clinical trial results does not depend on the similarity of the trial and reference populations. Rather it depends on having access to data that enable a reasonable exploration of the factors that are associated with variation in the magnitude of the treatment effect [11].

The second criticism of the reliance on randomized trials is that the costs are protocol driven and do not reflect the true costs of therapy when used in the community [24]. For example, participants are likely to have a fixed number of visits to a physician, more tests than are likely in ordinary clinical practice, and are likely to receive medications for adverse events that may not have been reported by patients had they not been in the trial. It is argued that observational studies provided a more accurate picture of the true costs of treatment with drugs in the community.

It is true that protocol costs are unlikely to reflect costs

of using the drug in the community and that some extrapolation from trial results will be required to give an indication of cost differences between two drugs. However, it does not follow that observational studies will provide better estimates of such a differential than randomized trials combined with informed modelling. Consider the example of the results of a randomized trial of drug and an active comparator. It is possible to identify the costs that are likely to differ between the two drugs in community use, based on the results of the randomized trial. While certain costs will relate to the design of the trial, for example monitoring and tests, other costs will reflect real differences between the two drugs, for example, differences in prices, in the rate of adverse events and their treatment, and differences in concomitant medication. And finally, certain outcome measures will indicate likely differences in costs of the two drugs when used in the community. For example, if only one drug causes abnormalities of liver function, then only patients using that drug will require monitoring of liver function tests in practice.

The additional information required to extrapolate from a randomized trial is likely to be similar to the additional information required by an observational study: e.g. prescribing patterns and compliance in the community setting. However, in our view, the randomized trial provides a better basis for such an extrapolation, particularly when the cost of two drugs are being compared. It would be reasonable to compare the pre-listing estimates of use and outcomes with actual results after listing. Listings are regularly reviewed, but to be comprehensive, this comparison requires sources of data that are not yet available in Australia.

Basing estimates of efficacy on results of observational studies

Observational studies can be an important part of an economic analysis, but they are not usually a key part of the assessment of comparative effectiveness. They are often the best source of data available to estimate the likely prevalence of use of a new drug, or patterns of practice behaviour that may influence overall resource use. Occasionally, in situations where there are no randomized trials, they may be the only data with which to assess efficacy. However, the main problem with observational data is the difficulty in estimating the size of the beneficial effect of a new drug. Observational studies may provide highly inaccurate estimates of treatment effects because of confounding. The main purpose of randomization is to eliminate selection bias. Due to the play of chance, randomization often leads to treatment and control groups that are comparable in respect of major known confounders but this is not its main purpose. In observational studies known differences between exposed and non-exposed groups can be minimised by matching, or statistical adjustment, but selection bias and residual confounding remain major threats to the validity of the study. Patients and prescribers choose drugs for a variety of reasons that may influence the outcome of treatment. These influences may be subtle, related to the indications for treatment or patient characteristics and may not be reflected in variables that are recorded in the study, particularly if there is reliance on existing databases. When comparing actual treatment alternatives in an economic analysis the true differences between the treatments may be smaller than the effects of residual confounding. Confounding may lead to substantial over- or underestimation of efficacy and may also influence the pattern of resource consumption by a cohort of users of a new drug.

Sheldon [6] cites a number of examples, based on published studies, where there has been a substantial difference between cost-effectiveness estimates derived from randomized trials and those derived from observational studies. Cholesterol-lowering drugs and selective serotonin reuptake inhibitors in particular were suggested to be less cost-effective when the analyses were based on the results of clinical trials, rather than the combination of observational data and models. This is consistent with the Australian experience.

Recommendations

The Australian experience to date has reinforced the original view that wherever possible, randomized trials must be the basis for determining differences between the estimates of efficacy and effectiveness of different drugs. In rare situations, it truly may not be possible to conduct randomized trials [17], but in general, observational studies cannot be the basis for estimates of effect.

As discussed by Freemantle & Drummond [18], the economic evaluation of pharmaceuticals will be enhanced by the availability of large simple trials, with relevant outcome measures and resource consumption documented carefully. We would add that increasingly, these will need to be active comparator trials, so that it will be possible to accurately compare the new treatment(s) with the old. It should be possible to avoid the sort of problem exemplified by the new anticonvulsant drugs, where despite the sequential development of lamotrigine, gabapentin, vigabatrin, topiramate and zonisamide, there are no trials that directly compare any two of the drugs [7].

From the clinical standpoint, we particularly need to know about small differences between the effects of drugs that are used for the treatment of common conditions. For example, small differences between the effects of new antidepressants and anticonvulsants may have significant influences on patient outcomes and a large impact on the public health outcomes. However, current clinical trials are not designed to detect these differences, although it is likely that the demand for this information will increase, from both purchasers, clinicians and those concerned with the public health aspects of common diseases.

In order to satisfy the need for accurate estimates of differences between treatments, it may be necessary to reconsider some aspects of the current standards for the conduct and design of clinical trials. If we need large trials, must they also conform to all the data requirements for trials for drug regulatory authorities? If the condition of interest has major public health implications, should commercial interests be expected to completely cover the cost of the trial? It has been argued that such trials could add years to the drug development process. A way forward that has been suggested may be for governments to co-sponsor clinical

trials in key public health areas [19], and for scientists and drug developers to consider early in drug development how clinical trials might best be conducted to measure clinically and economically important differences.

References

1 Drummond M. Basing Prescription Drug payment on Economic Analysis: The Case of Australia. *Health Affairs* 1992; Winter: 191–206.

2 Langley PC. The November 1995 Revised Australian Guidelines for the Economic evaluation of Pharmaceuticals. *Pharmacoeconomics* 1996; **9**: 341–352.

3 Commonwealth Department of Human Services and Health. *Guidelines for the Pharmaceutical Industry on Preparation of Submissions to the Pharmaceutical Benefits Advisory Committee*. 1995 Australian Government Publishing Service, Canberra.

4 Grobler MP, Macarounas-Kirchmann K, Pearce GA, Stafford M. Industry comment on the 1995 revised Australian Pharmacoeconomic Guidelines. *Pharmacoeconomics* 1996; **4**: 353–356.

5 Laupacis A, Feeny D, Detsky AS, Tugwell PX. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. *Can J Med Assoc* 1992; **146**: 473–481.

6 Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. *Br Med J* 1996; **313**: 36–39.

7 Marson AG, Kadir ZA, Chadwick DW. New antiepileptic drugs: a systematic review of their efficacy and tolerability. *Br Med J* 1996; **313**: 1169–1174.

8 George SL. Reducing patient eligibility criteria in cancer clinical trials. *J Clin Oncol* 1996; **14**: 1364–1370.

9 Remington RD. Potential impact of exclusion criteria on results of hypertension trials. *Hypertension* 1989; **13** (Suppl I): I-66–8.

10 Rapaport MH, Frevert T, Babior S, Zisook S, Judd LL. A comparison of demographic variables, symptom profiles, and measurements of functioning in symptomatic volunteers and an outpatient clinical population. *Psychopharmacology Bulletin* 1995; **31**: 111–114.

11 Rothwell PM. Can overall results of clinical trials be applied to all patients? *Lancet* 1995; **345**: 1616–1619.

12 Schmid CH, McIntosh M, Cappelleri JC, Lau J, Chalmers TC. Measuring the impact of the control rate in meta-analysis of clinical trials. *Controlled Clin Trials* 1995; **16** (Suppl): 66S.

13 Collins R, Peto R, MacMahon S, *et al*. Blood pressure, stroke, and coronary heart disease. Part 2. Short-term reductions in blood pressure: overview of randomized drug trials in their epidemiologic context. *Lancet* 1990; **335**: 827–838.

14 Glasziou PP, Irwin LM. An evidence based approach to individualising treatment. *Br Med J* 1995; **311**: 1356–1359.

15 Sackett DL. Applying overviews and meta-analyses at the bedside. *J Clin Epidemiol* 1995; **48**: 61–66.

16 Sheldon TA. Problems of using modelling in the economic evaluation of health care. *Health Economics* 1996; **5**: 1–11.

17 Walker AM. Common language. *Pharmacoepidemiology and Drug Safety*. 1996; **5**: 415–418.

18 Freemantle N, Drummond M. Should clinical trials with concurrent economic analyses be blinded? *J Am Med Ass* 1997; **277**: 63–64.

19 Glasziou PP. Support for trials of promising medications through the Pharmaceutical Benefits Scheme. *Med J Aust* 1995; **162**: 33–36.

(*Received 18 March 1997, accepted 17 June 1997*)

