Problems With the Interpretation of Pharmacoeconomic Analyses
A Review of Submissions to the Australian Pharmaceutical Benefits Scheme

Suzanne R. Hill, PhD, FAFPHM
Andrew S. Mitchell, MMed Sci
David A. Henry, MBChB, FRCP

Context  Pharmacoeconomic analyses are being used increasingly as the basis for reimbursement of the costs of new drugs. Reports of these analyses are often published in peer-reviewed journals. However, the analyses are complex and difficult to evaluate.

Objective  To describe the nature of problems encountered in the evaluation and interpretation of pharmacoeconomic analyses used as a basis for reimbursement decisions.

Data Sources  All major submissions to the Department of Health and Aged Care (DHAC) by the pharmaceutical industry for funding made under the Australian Pharmaceutical Benefits Scheme. Specifically, the DHAC’s database of submissions that were received between January 1994 and December 1997 were reviewed.

Study Selection  Of a total of 326 submissions, 218 had serious problems of interpretation and were included in the analysis. The nature of the serious problems reviewed were classified as estimates of comparative clinical efficacy, comparator issues, modeling issues, and calculation errors.

Data Extraction  All submissions in the DHAC’s database were reviewed and data were extracted if both the DHAC evaluators and technical subcommittee considered problems to have a significant bearing on the decisions of the parent committee.

Data Synthesis  Of a total of 326 submissions, 218 (67%) had significant problems and 31 had more than 1 problem. Of the 249 problems identified, 154 (62%) related to uncertainty in the estimates of comparative clinical efficacy, and 71 (28.5%) related to modeling issues, which included clinical assumptions or cost estimates, used in the construction of the economic models. There were 15 instances of disagreement over the choice of comparator, and serious calculation errors were found on 9 occasions. Overall, 159 problems (64%) were considered to be avoidable.

Conclusions  Significant problems were identified in these pharmacoeconomic analyses. The intensive evaluation process used in the Australian Pharmaceutical Benefits Scheme allowed for identification and correction of pharmacoeconomic analysis problems, but the resources that are required may be beyond the capacity of many organizations, including peer-reviewed journals.

JAMA. 2000;283:2116-2121  www.jama.com

For editorial comment see p 2158.
encountered during the evaluation and interpretation of the submissions evaluated between 1994 and 1997. The intention of this article is to discuss the implications of these findings for decision makers and journal editors who have to deal routinely with pharmacoeconomic analyses.

METHODS

The Australian Pharmaceutical Benefits Scheme and the process for evaluating drugs have been described previously. Subsidization of prescription drugs for use in the community (excluding public hospitals) is a Commonwealth (federal) government function. The Pharmaceutical Benefits Scheme is a comprehensive, publicly funded insurance program that reimburses pharmacists for the costs of a selected range of prescription drugs. There is a system of co-payments, and drugs are placed in different categories of access, based on evidence of their comparative effectiveness and cost-effectiveness in defined patient groups. Decisions to place new drugs in the Pharmaceutical Benefits Scheme are made by the federal health minister on the advice of a statutory committee, the Pharmaceutical Benefits Advisory Committee (PBAC). This is made up of family medicine practitioners, specialist physicians, clinical pharmacologists, pharmacists, and a consumer representative. The PBAC receives advice from a technical economics subcommittee comprising individuals with expertise in the fields of health economics, decision analysis, clinical epidemiology, and biostatistics.

Pharmaceutical companies make submissions in support of listing new drugs (or altered indications for existing drugs) to the Department of Health and Aged Care (DHAC). The companies follow guidelines issued by the DHAC, which lay out a detailed format for presenting the necessary data. Analysts, who may be employees of the company or external consultants, prepare the submissions.

Each submission is subjected to detailed evaluation by staff at the DHAC and their consultants. Typically, such an evaluation takes up to 2 person-weeks and involves checking the literature search used in compiling the submission, verification of trial results, validation of key assumptions in models, and confirmation of resource costs according to a manual of Australian costs. Evaluators rerun literature searches, check original clinical sources, and frequently run computer models that are provided by the sponsors. Members of the technical subcommittee of PBAC review the sponsor’s submission and the departmental evaluation. The subcommittee produces a summary document outlining key issues and the implications these have for recommendations made by the parent committee. The PBAC considers the sponsor’s submission, evaluation, report from the subcommittee, sponsor’s response to the evaluation, and views of its members when making its final recommendation to the federal health minister.

The DHAC maintains a database of all applications evaluated since 1994 (Table 1). We reviewed submissions received between January 1994 and December 1997.

We regarded problems in the submissions as “significant” if both the evaluators and technical subcommittee considered that the problem could have a significant bearing on the decisions of the parent committee. For instance, a positive recommendation for listing a new drug might be negated if the supporting data did not provide persuasive evidence of superior efficacy (compared with the established treatment), or where correction of a spreadsheet error made a cost-effectiveness ratio unattractive. To categorize the problems we used an article by O’Brien describing some of the difficulties encountered during the conduct of pharmacoeconomic analyses. O’Brien listed 7 serious issues, and we modified these to produce the categorization given in Table 2.

The database of submissions and the committee reports were scrutinized by 2 of authors (S.R.H. and D.A.H.). If a problem was noted in the committee report, it was categorized according to Table 2. Differences of opinion between the investigators were resolved by consensus.

Where appropriate, confidence intervals (CIs) for the proportions were calculated using the normal approximation.

In compiling this review we were bound by the secrecy provisions of the Australian National Health Act (1953, § 134A). The secrecy provisions of the act prevent the publication of detailed information that might reveal an individual drug, so we are able to identify drugs only by their clinical indications.

RESULTS

In the period 1994-1997, the PBAC reviewed 326 major applications. Of these, 182 were applications for new listings on the Pharmaceutical Benefits Scheme, and 51 were applications for major changes in indications, conditions of use, or prices applying to drugs that were already listed. The remainder were resubmissions (where a submission has been rejected previously) or a review of the basis of its pricing negotiations requested by manufacturers.

©2000 American Medical Association. All rights reserved.
Overall, 159 problems (64%; 95% CI, 68%-78%) were identified; the numbers in each category are summarized in Table 2. Of 326 submissions, 218 (67% [95% CI, 62%-72%]) were considered to have been avoidable during the planning and conduct of the pharmacoeconomic analysis.

### Table 2. Categorization of Types of Problems Identified in Pharmacoeconomic Analyses Submitted (N = 218) to the Australian PBS, 1994-1997*  

<table>
<thead>
<tr>
<th>Nature of Problems</th>
<th>Details of Problems</th>
<th>No. (%) of Problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate of comparative clinical efficacy (n = 154)</td>
<td>Availability of trials</td>
<td>12 (4.8)</td>
</tr>
<tr>
<td></td>
<td>Poor-quality trials</td>
<td>31 (12.4)</td>
</tr>
<tr>
<td></td>
<td>Analysis of interpretation of trial results</td>
<td>32 (12.9)</td>
</tr>
<tr>
<td></td>
<td>Use of surrogate outcomes</td>
<td>15 (6.0)</td>
</tr>
<tr>
<td></td>
<td>Determining therapeutic equivalence</td>
<td>64 (25.7)</td>
</tr>
<tr>
<td>Comparator issues (n = 15)</td>
<td>Uncertainty about choice of comparator or inappropriate comparator</td>
<td>15 (6.0)</td>
</tr>
<tr>
<td>Modeling issues (n = 71)</td>
<td>Technical aspects of the model</td>
<td>24 (9.6)</td>
</tr>
<tr>
<td></td>
<td>Unsubstantiated assumptions</td>
<td>15 (6.0)</td>
</tr>
<tr>
<td></td>
<td>Uncertainty about costs</td>
<td>32 (12.9)</td>
</tr>
<tr>
<td>Calculation errors (n = 9)</td>
<td>Errors that introduced serious inaccuracies in the estimation of the cost-effectiveness ratios</td>
<td>9 (3.6)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>249 (100)</td>
</tr>
</tbody>
</table>

*PBS indicates Pharmaceutical Benefits Scheme. Some submissions had more than 1 problem. Percentages do not sum to 100 because of rounding.

### Sources of Clinical Data

Of 326 submissions, 279 (86% [95% CI, 82%-89%]) contained economic analyses based on the results of randomized trials. A total of 238 submissions (73% [95% CI, 68%-78%]) contained head-to-head trial comparisons of the new agent and the chosen comparator. In 41 cases the comparisons between the new drug and comparator were made by considering the results of 2 sets of trials with a common reference, usually placebo. Twenty-six submissions were based on quasi-experimental designs, and 21 were based on uncontrolled data. In all, 64 submissions (19.6%; 95% CI, 15%-24%) based their estimates of comparative clinical effectiveness on the results of meta-analyses. Of these, 17 were published meta-analyses; during the last year of this survey (1997), 2 of these were drawn from the Cochrane Library.

### Overview of the Database

Of 326 submissions, 218 (67% [95% CI, 62%-72%]) were considered to have presented serious problems of interpretation. Thirty-one submissions had more than 1 problem. The proportion of submissions with problems did not change over time: 69.7% in 1994, 69.2% in 1995, 63.1% in 1996, and 65.9% in 1997. A total of 249 serious problems were identified; the numbers in each category are summarized in Table 2. Overall, 159 problems (64%; 95% CI, 58%-70%) were considered to have been avoidable during the planning and conduct of the pharmacoeconomic analysis.

### Problems With Estimates of Comparative Clinical Efficacy

Problems in this category reflected uncertainty about the magnitude of any clinical benefit of the new drug compared with existing agents. The weight given to this reflects the priority the PBAC gives to establishing the comparative clinical performance of new drugs before making a judgement regarding their economic performance.

### Availability of Trials

In a number of submissions, including those relating to drugs for Parkinson disease, breast cancer, and ovarian cancer, no randomized trial was available, and data from uncontrolled studies were used to estimate comparative cost-effectiveness. The magnitude of benefit of each drug had to be inferred from uncontrolled comparisons of series of patients receiving the drug of interest or the comparator. In some cases, data from a case series involving the new agent were compared with the outcomes from the single arm of a randomized controlled trial involving the comparator.

In a further group that included drugs for the treatment of chemotherapy-induced vomiting, glaucoma, asthma, intermittent claudication, and bacterial vaginosis, difficulties arose because of incomplete presentation of potentially relevant trials. Literature searches performed by the evaluators identified randomized trials that were relevant to the clinical comparison in the economic analyses, but had not been mentioned by the sponsor. The data from these trials modified, and in some cases contradicted, the claims made by the sponsor.

### Poor-Quality Trials

In 31 submissions that included treatments for cancer, insomnia, osteoporosis, and osteoarthritis, the key clinical trials had serious methodological flaws. For example, a submission for a new drug claimed that it had a lower rate of adverse effects, but this was based on an open-label trial with an inadequate sample size and unblinded assessment of subjective outcomes.

### Analysis and Interpretation of Trial Results

The statistical analyses carried out in submissions were often complex and involved a range of meta-analytic techniques, subgroup analyses, or reanalyses of clinical trial data. Problems included inappropriate subgroup analysis, inappropriate adjustment of event rates, and problems with statistical pooling of data.

Approximately 20% of submissions used meta-analyses of the results of several randomized trials to estimate the comparative clinical benefit of the new drug. In 3 instances, involving drugs for diabetes, osteoporosis, and human immunodeficiency virus (HIV) infection, difficulties were encountered when interpreting pooled data from crossover trials.

In a group of submissions including drugs to treat psychosis and eradicate Helicobacter pylori, problems arose while linking the evidence from the randomized trials to the proposed indication. This was because the proposed use was for “second-line” treatment (at a higher proposed price), whereas the trials had been conducted in a “first-line” setting. It was not clear that patients with more severe or “resistant” disease would respond in the same way.
as those who had been included in the randomized trials.

A number of submissions, including drugs for cancer, asthma, prevention of thromboembolic disease, and treatment of respiratory tract infections, were based on claims that the new products were superior to the comparators. The trials were considered to be of a satisfactory standard, but evaluators disagreed with the sponsors’ claims. There were 2 common explanations—the trials were insufficiently powered to show differences between 2 active treatments, or the demonstrated differences in outcomes were considered to be clinically insignificant.

Use of Surrogate Outcomes. Evaluation was sometimes hampered by reliance on surrogate outcomes. Examples included submissions for drugs for dementia that used measures of cognitive function rather than measures of social coping; drugs for Paget disease of bone, where biochemical tests were used rather than symptoms or measures of disability; and urinary flow rates and lower urinary tract symptom scores in men undergoing treatment for benign prostatic hypertrophy, rather than reduction in surgery or prevention of progression to renal failure.

Determining Therapeutic and Dose Equivalence. The most frequent problem involved determining whether the available data for a new product supported the manufacturer’s claim that the product was therapeutically equivalent to a comparator. There were also difficulties in determining therapeutically equivalent doses. The latter is important during price setting by cost minimization analysis.

In many cases, including drugs for hypercholesterolemia, Paget disease, and hormone replacement therapy, the uncertainty was because trials were too small to exclude clinically significant differences or were conducted over too short a time. In assessing dose equivalence, difficulties arose because of the design of the comparative trials, for example, a comparison of fixed doses of 1 product with variable doses of the comparator.

Choice of Comparator
The Australian guidelines request sponsors to use as comparator the “treatment most likely to be replaced” by the new therapy. There were 15 examples, involving drugs for Parkinson disease, epilepsy, infectious diseases, and osteoporosis, in which there were disagreements regarding the correct choice of comparator. In some cases, no comparator was nominated, and no comparative data were provided. In other cases, the comparator that was nominated was the least prescribed and most expensive alternative.

When there are no randomized trials that directly compare the new drug and the comparator, the Australian guidelines advise companies to base their submission on the results of an indirect comparison. This involves 2 sets of trials, with placebo or another active drug as a common reference. Assessing equivalence or superiority in this situation has proved difficult because of confounding at study level and variation in the baseline severity of disease in the control groups of the respective trials.

Examples of this type of problem include anticonvulsant drugs and drugs for Parkinson disease and HIV infection.

Modeling Issues
Technical Aspects of Model. These problems involved examples such as discounting costs but not benefits (discounting is the technique used in economic evaluation to calculate the present values of costs and consequences of an intervention, given that the effects of the intervention may occur in the future rather than the present13), failing to relate costs and outcomes appropriately, and uncertainties arising from extrapolating benefits seen in a short-term trial over the lifetime of the patient. More recent submissions have used models based on willingness-to-pay studies and time-trade-off analyses. Problems identified in this group in particular have included inappropriate questionnaire design and inadequate sample size.

Unsubstantiated Assumptions. Two main issues emerged in this group. In submissions that involved preventive treatments in osteoporosis, hypertension, and hypercholesterolemia the models provided estimates of benefit that were biologically implausible and unsupported by controlled clinical data. The second problem was the derivation of utility estimates from insignificant or uncertain clinical data. This was evident in several submissions for cancer therapies and also for drugs used in the treatment of symptoms of relatively benign conditions, such as minor infections.

Estimation and Incorporation of Costs. In some instances, cost offsets were unduly optimistic, without supporting data, or benefits and associated cost savings were overestimated because of inappropriate or truncated analyses. Lack of transparency in the calculation of costs and outcomes was a significant problem, despite the fact that the submissions usually contained copies of spreadsheets or models.

Calculation Errors
These problems included failure to calculate an incremental ratio correctly, unbalanced numbers of patients assigned to alternative treatments in a model, and simple spreadsheet errors that resulted in a product being erroneously portrayed as dominant.

COMMENT
Establishing a formal link between measures of costs and outcomes of drug therapy is an appealing approach to drug purchasing decisions. This comes closer to the notion of a true “market” than alternatives, such as leaving manufacturers to obtain the highest prices from poorly informed purchasers or government control of profits as practiced in the United Kingdom.15 The variations seen in the prices of very similar drugs in many countries16 reflect a degree of market failure. This arises when purchasers or insurers use imperfect tools to assess the community’s needs, the extent to which a range of drugs meets these needs, and at what price. There are strong arguments for linking purchasing and pricing decisions to the therapeutic perfor-
mance of drugs; this is the central argument for the use of pharmacoeconomic analysis. Our experience indicates that even when operating under a prescriptive regime, with clear guidelines underpinned by strong legislation, problems frequently arose during the conduct and interpretation of pharmacoeconomic analyses.

The bulk of problems concerned the interpretation of comparative clinical data. Drug development programs mainly are designed to meet the needs of drug regulatory authorities. Standards for evaluating efficacy, toxicity, and manufacturing quality are well established.\textsuperscript{17,18} These represent an attempt to balance the need for adequate evidence on efficacy, safety, and quality with the desire of the community for rapid access to important new drugs and the commercial interest of manufacturers in obtaining licensing approval as soon as possible. Pharmacoeconomic data are used in an environment that is competitive and market driven. The desire of manufacturers is to show new products in the most favorable light and establish an advantage over their competitors. Unfortunately, the highest-quality data are sometimes suboptimal for performing analyses of comparative efficacy and costs.\textsuperscript{19} The Australian guidelines provide a rationale for choosing the most appropriate comparator when conducting pharmacoeconomic analyses. Typically, it is the drug most likely to be replaced by a new product. This depends on local factors, and high-quality comparative trials of the new drug and this comparator may not be available. The unsuitability of the available data and the desire of companies to get their drugs to market sometimes lead to claims of superiority over competitors that are not supported after close examination of the clinical data.

Our impression is that analysts involved in the design and conduct of pharmacoeconomic analyses can have difficulty substantiating claims made by companies regarding the comparative clinical performance of their products. We found that estimates of clinical performance were sometimes based on uncontrolled studies, including case series, comparisons of single arms from different trials, or inadequately conducted comparative trials. Companies often relied on indirect comparisons, made through 2 sets of trials with a common comparator (eg, placebo). The Australian guidelines encourage the use of such evidence when direct comparative trials are unavailable; however, estimates of comparative performance are highly likely to be confounded by factors that are unequally distributed between the different study populations. Flawed estimates of comparative clinical performance were sometimes used to justify a claim for higher prices. Some of these claims violated well-established rules of epidemiological and statistical inference.

We believe that our experiences revealed no basic intention to deceive. The occasional failure to present relevant trials became less common after the guidelines made full disclosure a requirement. Where there were errors, for instance simple numerical miscalculations, or mistakes in transcribing probabilities into a decision tree, they were readily detected by examination of the spreadsheets and probably reflected inadequate quality control. Most of the problems related to interpretation of clinical data. It seemed that company employees had optimistic views of their product's performance, and analysts had to make do with suboptimal and poorly designed studies when making inferences regarding comparative clinical performance. Complex modeling techniques do not overcome fundamental deficiencies in clinical data.

Despite these concerns, which have been voiced in other review articles,\textsuperscript{6,8,20} the use of economic analyses as an aid to making decisions about allocating resources in health care is increasing. A number of jurisdictions now have published guidelines for economic analysis, and organizations such as health maintenance organizations and national and provincial governments are considering results of analyses when making assessments about new pharmaceuticals and health technologies.\textsuperscript{21,22}

In light of our experience, do we support the widespread use of pharmacoconomics in decision making? Overall our answer is yes, but this requires important qualification. Clearly, there are many methodologic issues that need to be considered in the evaluation of these analyses. Initially, we were surprised at the number of submissions that had significant methodologic problems. Over time we formed the view that this was not unexpected given the nature of the process, which requires a complicated synthesis of data from a variety of sources. The problems we report here had the potential to distort the estimate of the cost-effectiveness ratio for each product and therefore the decisions that were based on the interpretation of this parameter. It should be noted that other factors are considered in the decision-making process, such as clinical need, equity of access, “rule of rescue” (\textit{rule of rescue} is the desire of most societies to spend large amounts of money to save individuals who are in extreme danger; applied to drugs, it means accepting some cost-ineffective interventions for patients with rare catastrophic illnesses who have no other treatment options), and the total cost to the health care system, so that a flawed evaluation did not necessarily lead to the drug being rejected for subsidy.

The evaluation process that identified these problems was demanding and required (typically) close examination of the data by several staff members with training in clinical epidemiology and health economics. Sometimes identification of key issues required examination and reanalysis of source data; sometimes it required examination of a spreadsheet. Because we were working within a national program for reimbursement of pharmaceuticals (annual expenditure >A$3 billion) it was possible to direct the resources needed to evaluate these analyses fully. In many cases, the analyses were corrected to allow a recommendation about reimbursement to be made by the parent committee.
PROBLEMS WITH PHARMACOECONOMIC ANALYSES

In our view, any agency or organization that wishes to make formal use of pharmacoeconomic data must appreciate the intensity of the evaluation process that is necessary to ensure that decisions are based on accurate data. We doubt whether any conventional peer-review process is adequate. Such concerns have been expressed by others, and a recently published article and editorial have pointed to possible bias in the qualitative conclusions in pharmacoeconomic analyses published by academic investigators.7,23-26 Published pharmacoeconomic analyses are subject to the same constraints as any scientific study. They have to comply with journal space requirements that limit the presentation of relevant data. Complex models and calculations and details of assumptions and methods have to be truncated to comply with journal and editorial requirements. Referees are seldom provided with the computer models, and most will not have the time or inclination to run them. A 10-page manuscript is in stark contrast to the 2- and 3-volume (400-page) submissions containing the details of the economic evaluations that are routinely evaluated for the Australian Pharmaceutical Benefits Scheme.

As mentioned earlier, we do not believe that the problems identified in the review were deliberately introduced. Most of the problems were the result of the intrinsically complex nature of economic analyses and genuine differences of opinion about how to interpret the results of clinical trials. This makes it important for full details of analyses to be available to reviewers, readers, and decision makers, so that they can make informed judgments regarding analysis validity. In our view, conventional publication in peer-reviewed journals is not a realistic means of meeting this requirement.

Funding/Support: This work was supported by a grant from the Commonwealth Department of Health and Aged Care.

REFERENCES


©2000 American Medical Association. All rights reserved.