No 36 April 1998

OHE briefing

TRENDS IN ECONOMIC EVALUATION

CLIVE PRITCHARD

1. INTRODUCTION

The aim of this paper is to consider recent trends in the volume and characteristics of economic evaluation literature by interrogating the Health Economic Evaluations Database (HEED), and to offer some comments on how the quality of such studies can be assessed by reviewing existing work which has considered this issue. HEED has been developed as a joint initiative between the Office of Health Economics (OHE) and the International Federation of Pharmaceutical Manufacturers' Associations (IFPMA). The main aim of the database is to provide structured summaries (reviews) of articles appearing in the literature relevant to the economic assessment of health technologies including articles which are themselves reviews of the literature. The database also includes, in bibliographic detail, entries from existing databases of economic evaluation literature, as discussed below.

The structure of the paper is as follows. Section two provides an overview of the data included on HEED and the way in which it is compiled. Section three presents information on the growth in literature, including the total numbers of references, and gives a breakdown by type of entry, for example the numbers of applied studies as compared with the numbers of reviews of applied studies. Applied economic evaluation studies are those which make an original attempt to bring together information on costs and outcomes, and include not only those studies which are based on the collection of original (primary) clinical or cost data, but also those which rely on the adaptation of secondary data (literature) sources, such as modelling studies. Other applied studies which are not economic evaluations (cost of illness studies and cost analyses) incorporate an original element of cost estimation. The distribution by type of economic evaluation (e.g. cost effectiveness analysis, cost utility analysis) across applied studies is also presented. The glossary of terms in Box 1 gives definitions of the different types of evaluation considered relevant to be included in HEED.

Section four concentrates exclusively on examining a number of aspects of applied studies. Firstly, in order to give an idea of the subject matter covered by these studies, the spread across disease areas will be illustrated by comparing the distribution for 1992 and 1996 over the chapters of the International Classification of Diseases (Clinical Modifications), ninth revision (ICD-9) classification system for 1992 and 1996. Similar comparisons are presented for the distribution of pharmaceutical evaluations across the chapters of the Anatomical Therapeutic Chemical (ATC) classification scheme and for all applied studies across types of technology (pharmaceutical, surgical, screening etc.). The distinction between pharmaceuticals and other types of technology is examined in comparing the distribution of study by type of sponsor and by study design, focusing

Contents

Box 1

The following types of analysis include information both on the resource costs and health outcomes of an intervention:

Cost minimization analysis (CMA). In this form of analysis, the outcomes of two or more interventions being compared are taken to be identical, perhaps on the basis of previously published results or original data which provides no evidence of a difference in outcomes, and the interventions in question are compared on the basis of their relative costs.

Cost effectiveness analysis (CEA). This form of analysis focuses on a single outcome measure, often one specific to the disease in question, with the results expressed in terms of cost per unit of health outcome, e.g. cost per 10% reduction in LDL cholesterol level, cost per case detected, cost per life year gained.

Cost utility analysis (CUA). This form of analysis uses an outcome measure which combines longevity and quality of life, usually the quality adjusted life year (QALY), with results expressed in terms of cost per QALY gained.

Cost benefit analysis (CBA). This form of analysis applies a monetary value to both the costs and the health outcomes of an intervention, with the results expressed as a net benefit (the difference between the monetary value of outcomes and costs) or a ratio of benefit to cost.

Cost consequences analysis (CCA). In this form of analysis, a range of outcome measures are presented alongside costs but no ratio of cost per unit of outcome is presented. Analyses defined as CCAs in HEED are frequently concerned with a single intervention without explicit consideration of an alternative; such studies may be considered partial rather than full evaluations and are sometimes labelled as 'cost-outcome descriptions' (Drummond et al., 1997).

The following types of analysis included on HEED are not forms of economic evaluation since they consider only the costs of an intervention:

Cost of illness. This form of analysis attempts to estimate the total costs associated with a disease (e.g. costs of treatment, costs of lost productivity), usually for a country, but sometimes limited to a smaller geographic area.

Cost analysis. This form of analysis compares the costs of treating a particular condition with one intervention versus another, with the results expressed in terms of a net resource saving or a ratio of resource savings to resource costs. Studies may be classified as both cost consequences analysis and cost analysis when outcomes are reported but the main focus of the study is to estimate the potential savings of one treatment relative to another.

Other terms used in this briefing:

Willingness to pay (WTP). The maximum monetary amount that is stated as an individual's willingness to sacrifice, for example in terms of increased taxes, for the provision of a service. This method is used to value health outcomes in cost benefit analysis.

Discounting. The opposite of compound interest, used to express costs and outcomes in future years in terms comparable to costs and outcomes experienced today. A rate of 5% per year is frequently applied in discounting calculations.

specifically on studies which can be classified as economic evaluations alongside randomized trials. Separate consideration is given to cost utility analyses to explore how utility measurement has been undertaken, while the final two sub-sections investigate the use of sensitivity analysis and discounting in all applied studies.

The topics covered in section four were selected on the basis either for comparison with the findings of previous studies charting trends in the economic evaluation literature or because it is a methodological issue of interest. For example, Mason and Drummond (1995) have described the distribution of entries in the Department of Health Register of Cost-Effectiveness Studies according to ICD-9 chapter while, in their economic evaluation bibliography, Elixhauser et al. (1993) discuss the types of technology assessed. Methodological aspects of the literature that have been addressed include the conduct of economic evaluations alongside clinical trials (Drummond and Davies, 1991), the quality of cost-utility analyses (Gerard, 1992) and the use of sensitivity analyses (Briggs and Sculpher, 1992). Although the paper does not attempt to provide an in-depth discussion of the quality of studies, the discussion should be of interest to readers of previous overviews of the economic evaluation literature, such as Warner and Hutton (1980) and Elixhauser et al. (1993).

Comments will be made in section five about the difficulties of making quality assessments and the usefulness of the typical checklist approach used by researchers when attempting to assess the quality of economic evaluations (e.g. Udvarhelyi, 1992). As part of this section, a number of studies assessing the quality of economic evaluations will be reviewed. Finally, section six will draw together some conclusions and suggest possibilities for further research.

2. THE HEALTH ECONOMIC EVALUATIONS DATABASE (HEED)

The HEED database, as of February 1998, contained approximately 14,000 references. There are two types of entry on the database: firstly, bibliographic entries and, secondly, references which have been reviewed according to a standard report format by a health economist. The latter constitute nearly 6,000 of the total of 14,000 references. Previous examples of sources of economic evaluation literature, in the form of the Battelle database (Elixhauser et al., 1993) and the Wellcome database (Backhouse et al., 1992), have been incorporated, in bibliographic form, into the HEED database. It will also shortly include, in reviewed form, those studies appearing in the Register of Cost-Effectiveness Studies (RCES) produced by the Department of Health (1994).

Like the NHS Economic Evaluation Database (NHS Centre for Reviews and Dissemination, 1996), the main objective of HEED is to add new references, in reviewed form, on an on-going basis. Articles to be reviewed are identified monthly from a number of data sources. The primary sources used are the following on-line databases: Medline, Embase, Health Planning and Administration, Psychinfo and DHSS-Data. Hand searching of a number of medical, health economics and policy journals is also undertaken and attempts are made to include all relevant grey literature. Database searches have covered the period from 1992 onwards.

Those articles which appear on HEED in reviewed format are categorised according to type of entry. These are 'applied' studies, 'reviews' of applied studies, 'methodological' studies (which deal with the overall methodology of economic evaluation, costing methodology or methods relating to utility measures or the monetary valuation of outcomes), and 'policy' papers, such as commentaries on the use of government guidelines for the conduct of economic evaluations. Letters, editorials and 'other' types of study also feature.

In terms of types of evaluation, there are three main categories. Firstly, there are those which satisfy the criterion for an economic evaluation of considering both the costs and health effects (outcomes) of an intervention. These can be classified according to the standard definitions of cost minimisation, cost effectiveness, cost utility and cost benefit analyses plus a fifth category of cost consequences analyses (which assess costs and outcomes but do not express the results in the form of a ratio). Secondly, cost of illness studies are included and, thirdly, there are cost analyses comparing the relative costs of two or more interventions for a given condition. These normally express their results in terms of a net cost or saving of one intervention relative to another or as a ratio of cost savings to cost. In order to illustrate the growth in literature appearing on the HEED database, the next section includes a presentation of the total number of references on the database and of the distribution of (reviewed) studies according to type of entry. For subsequent analyses, only applied studies have been included.

3. OVERALL GROWTH IN LITERATURE

Figure 1 shows the total number of database entries for each of the years 1992 to 1996. This indicates a similar level of activity for the years 1992, 1993 and 1994, but shows a large increase in activity for 1995. This is probably due in part to improvements in the coverage of the literature searches, although it is unlikely that all of the nearly doubling in numbers of studies could be explained by improvements in literature coverage. Based on a projection of the incomplete 1996 figures assuming that 1996 is approximately 95% complete[1], the level of activity appears to have been broadly maintained between 1995 and 1996.

Considering the distribution by type of entry in Figures 2 and 3, the most striking finding from these figures is the shift in the balance of type of entry. In particular, there has been a shift in emphasis towards applied studies which increased as a proportion of all reviewed articles from 42% in 1992 to 71% in 1996. Correspondingly, the proportion of reviews of applied studies has declined from 22% to 14% of reviewed articles and the proportion of methodological studies has fallen from 17% to 6%. The proportion of papers dealing with government and public policy matters has remained relatively low, at around 2% in 1992 and just over 1% in 1996.

Focusing specifically on applied studies, it is of interest, first of all, to examine the distribution of these studies by type of economic evaluation[2]. The most notable change in the distribution by type of evaluation is the significant increase in the importance of cost consequences analyses,

1. It is assumed that the 1996 data are approximately 95% complete on the basis of an estimate of the number of studies still to be retrieved for this year.

2. Unlike article type, type of economic evaluation is not a mutually exclusive categorisation since more than one type of evaluation may apply to any one study. Therefore, the figures shown in table one will sum to more than the total number of applied studies.

Figure 1 **Total references**

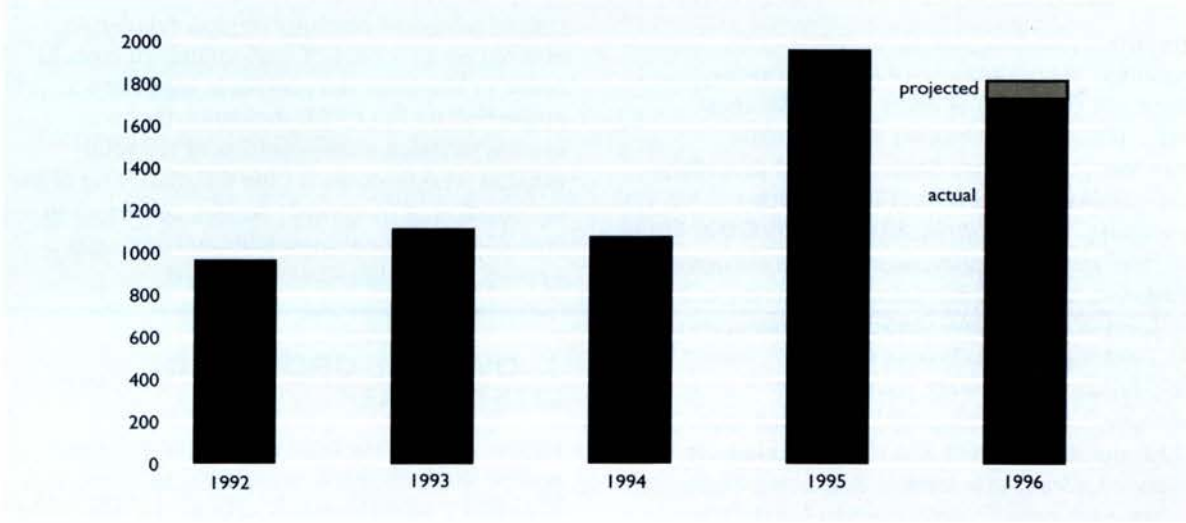


Figure 2 **Distribution of reviewed articles by study type 1992**

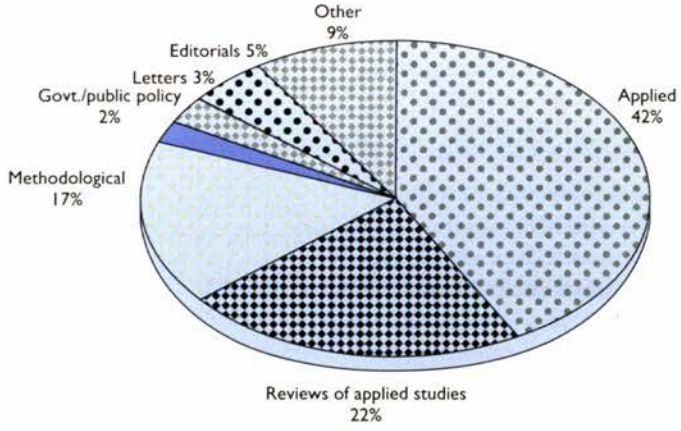
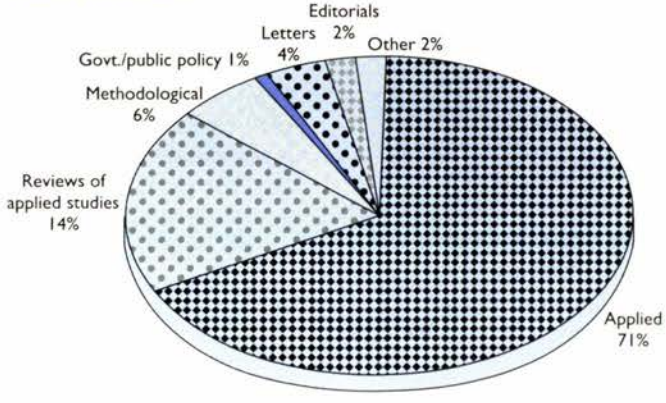


Figure 3 **Distribution of reviewed articles by study type 1996**



which present information on costs and outcomes but do not synthesis the two into a ratio. As indicated by Table 1, this type of analysis grew from 32% to 55% of applied studies between 1992 and 1996. Thus, the growth in literature seen across the period appears to have been most heavily influenced by applied cost consequences analyses. Alongside this change, the proportion of applied studies which are cost effectiveness analyses has fallen from 41% to 23%. Cost utility analyses continue to represent a relatively small proportion of studies, declining in importance from 9% to 5% with the number of cost benefit analyses (including willingness to pay) being negligible and, in 1996, being less than 1% of studies compared with around 5% in 1992. In absolute terms, the numbers of cost effectiveness and cost utility studies have experienced a four-fold increase while the number of cost benefit analyses has remained constant.

Table 1 **Types of Economic Evaluation - Percentages**

Types of Evaluation - Applied Studies (%)	1992	1993	1994	1995	1996
Cost minimisation	13%	10%	11%	11%	8%
Cost effectiveness	41%	40%	28%	20%	23%
Cost utility	9%	8%	5%	4%	5%
Cost benefit	5%	3%	3%	1%	0%
Cost consequences	32%	32%	40%	54%	55%
Cost of illness	5%	9%	7%	5%	5%
Cost analysis	19%	23%	22%	27%	19%
Total number of applied studies	111	163	398	1038	1053

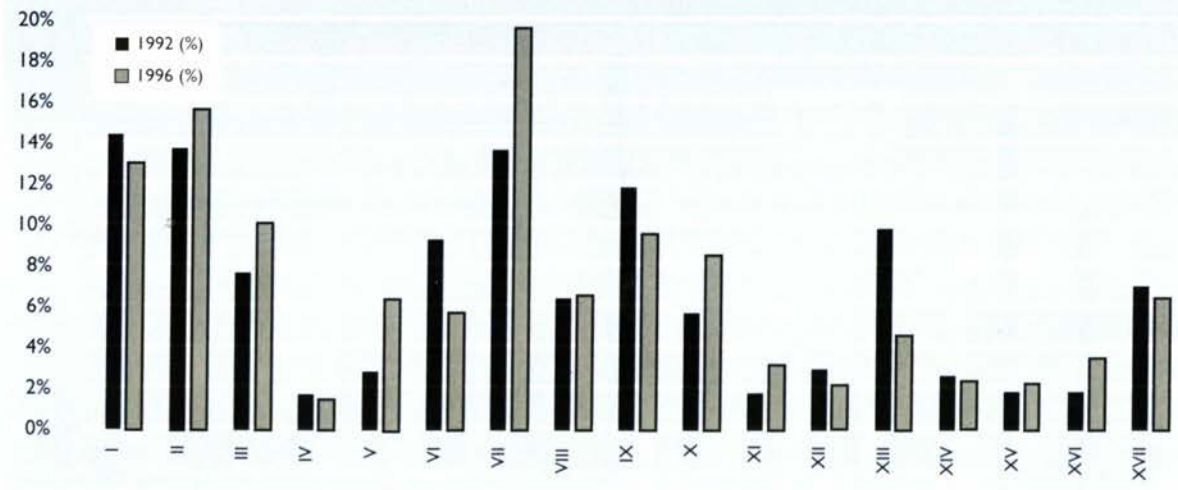
4. TRENDS IN LITERATURE

From the point of view of those interested in the issue of resource allocation decisions across programme areas, the paucity of cost utility studies is disappointing (although the use as an outcome measure of life years gained, which might be viewed as a second best measure as compared with QALYs, was observed in over a quarter of applied cost effectiveness analyses in 1996). However, economic evaluations can still be useful for comparing the cost-effectiveness of different interventions within therapeutic or disease areas. It is, therefore, of interest to investigate the distribution by disease area of studies contained within HEED. This can be done by analysing the distribution of entries across chapters of the International Classification of Diseases, ninth revision (ICD-9) classification system. This distribution will be presented in the next section, again comparing 1992 with 1996. We also examine the distribution of applied studies of pharmaceuticals (the most important technology assessed in terms of the proportion of studies undertaken) across the chapters of the Anatomical Therapeutic Chemical (ATC) classification system.

4.1 ICD-9 and ATC codes

Figure 4 shows the distribution of applied studies by chapter across the seventeen chapters of the ICD-9 classification system, comparing 1992 with 1996. Over the time period, there has been relatively little change in the distribution of studies by disease area. In 1996, the top three chapters in terms of the concentration of studies were chapter seven (diseases of the circulatory

Figure 4 **Distribution of applied studies by ICD-9 chapter**



I Infectious and Parasitic Diseases
II Neoplasms
III Endocrine, Nutritional and Metabolic Diseases, and Immunity Disorders
IV Diseases of the Blood and Blood-Forming Organs
V Mental Disorders
VI Diseases of the Nervous System and Sense Organs
VII Diseases of the Circulatory System
VIII Diseases of the Respiratory System
IX Diseases of the Digestive System
X Diseases of the Genitourinary System
XI Complications of Pregnancy, Childbirth, and the Puerperium
XII Diseases of the Skin and Subcutaneous Tissue
XIII Diseases of the Musculoskeletal System and Connective Tissue
XIIV Congenital Anomalies
XV Certain Conditions Origination in the Perinatal Period
XVI Symptoms, Signs, and Ill-Defiined Conditions
XVII Injury and Poisoning

system), chapter two (neoplasms) and chapter one (infectious and parasitic diseases); these were also the top three chapters in 1992.

The proportions of studies on diseases of the circulatory system and on neoplasms increased between 1992 and 1996, from 14% to 19%, while the proportions of studies classified under infectious and parasitic diseases (chapter one) and neoplasms (chapter two) were about the same. Diseases of the digestive system (chapter nine), the fourth most important chapter in 1992, had been overtaken by endocrine, nutritional and metabolic diseases, and immunity disorders (chapter three) by 1996. These figures show some similarities with those for the Wellcome database, in which diseases of the circulatory system, neoplasms and infectious and parasitic diseases were again the three most important individual disease areas. Diseases of the digestive system, however, appeared to be less important in the Wellcome database[3].

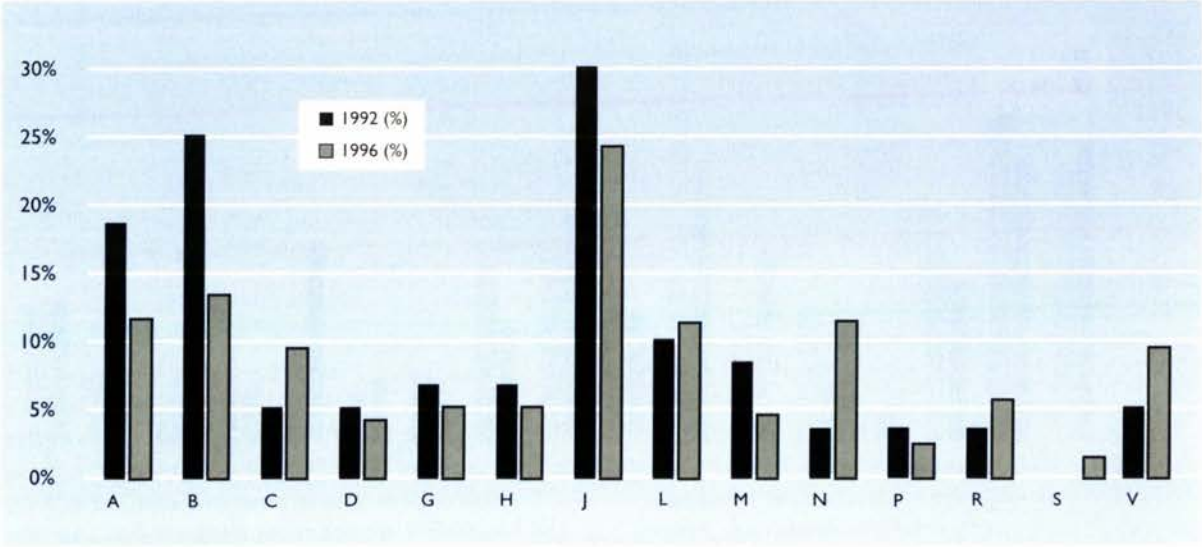
It is worth noting that two of the top three most significant disease areas for undertaking studies were the two disease areas found to have the greatest burden of illness in a study by Bowie et al. (1997). They found that, for the South West region of England in 1992, the burden of disease in terms of disability adjusted life years (DALYs) was greatest for diseases of the circulatory system and second highest for neoplasms when DALYs lost through premature mortality and those lost through disability were combined. However, other than these two instances, a close correspondence between the prevalence of studies and the burden of illness was not observed. For example, mental disorders account for a relatively small proportion of studies but were the third most important in terms of disease burden in the Bowie et al. (1997) study.

Within those ICD-9 chapters which have the greatest concentration of studies in HEED, there are particular diseases which appear more frequently than others. For example, around 20% of studies on neoplasms in 1996 dealt with breast cancer, while about 40% of studies on diseases of the circulatory system looked at ischaemic heart disease in this year. The sample sizes for 1992 are probably too small to make meaningful comparisons between 1992 and 1996 for particular diseases.

The distribution of all applied studies by ICD-9 chapter is reflected in the distribution of pharmaceutical evaluations by ATC chapter, in that there are again four or five chapters in which many studies are concentrated, with the remaining chapters having a single figure percentage share of studies (as set out in Figure 5). The top three chapters in both 1992 and 1996 are chapter J, General Anti-infectives for Systemic Use, chapter B, Blood and Blood Forming Organs and chapter A, Alimentary Tract

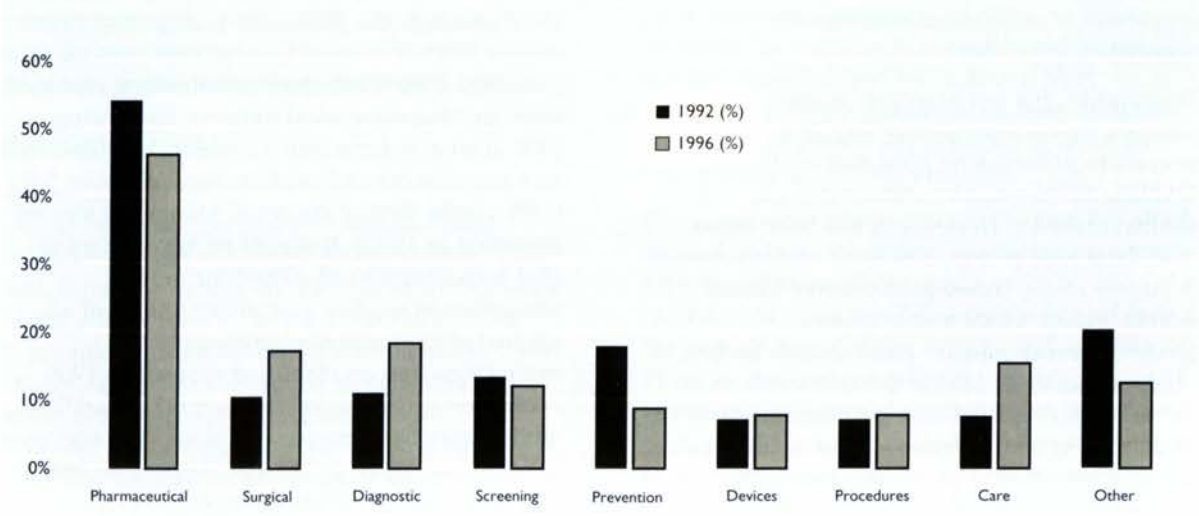
3. Since the Wellcome database has been incorporated into HEED, it might be expected that there would be a close similarity in the distribution of entries by disease area between the two. However, incorporation was conducted at the level of the bibliographic reference only. Therefore, the analysis of HEED presented here, of reviewed studies, excludes those included in the Wellcome database.

Figure 5 **Distribution of applied studies by ATC chapter**



A Alimentary Tract and Metabolism
B Blood and Blood Forming Organs
C Cardiovascular System
D Dermatologicals
G Genito Urinary System and Sex Hormones
H Systemic Hormonal Preparations, excluding Sex Hormones
J General Anti-Infectives for Systemic Use
L Antineoplastic and Immunomodulating Agents
M Musculo-Skeletal System
N Nervous System
P Antiparasitic Products, Insecticides and Repellants
R Respiratory System
S Sensory Organs
V Various

Figure 6 **Distribution of reviewed articles according to type of technology**



and Metabolism. The latter two chapters appear to have declined in importance between 1992 and 1996, with the proportion of drug evaluations in these chapters falling from 25% to 13% and from 18% to 12% respectively. The one chapter which increased its share substantially between 1992 and 1996 was chapter N, Central Nervous System. Within each of the top three chapters, particular sub-categories are more important than others. In chapter A, drugs for the treatment of peptic ulcer accounted for roughly 40% of studies in 1996, while antithrombotic agents accounted for over 50% of studies in chapter B. In chapter J, there was a greater spread of studies, although cephalosporins accounted for over 20% of studies in 1996. The numbers of studies for sub-categories within ATC chapters are too small to make meaningful comparisons.

The distribution of studies according to ATC chapter may be compared with drug sales by therapeutic category. According to IMS data reported in SCRIP (1997), cardiovascular drugs represent the largest therapeutic category by dollar sales for both Japan and the top seven European markets, but are second to central nervous system (CNS) drugs in the USA, in contrast with the relatively low share of studies for ATC chapter C. Anti-infectives, which have by far the largest share of studies, are only third or fourth most important in terms of sales for these three regions. In all three regions, alimentary/metabolism products are second most important in terms of sales, similar to their position in terms of studies undertaken. Comparisons such as these are, however, complicated by differences between region. For example, CNS drugs, which are fourth most important in terms of studies have a similar level of importance in sales to the top seven markets in Europe, in third place, while they are at the top of the sales league in North America, but are in seventh position in terms of sales for Japan. Just as there is no straightforward relationship between studies undertaken and cost or burden of illness according to ICD-9 code, neither is there one between studies carried out and drug sales according to therapeutic category.

4.2 Type of Technology Assessed

While the sales of pharmaceuticals and the absolute number of economic evaluations of pharmaceuticals continue to rise, it should be noted that the changes in numbers of studies by ATC chapter have taken place within a relative decline in the proportion of studies which include the evaluation of pharmaceuticals. This decline, in relation to other technologies, is shown in Figure 6. Those technologies for which a relative increase in economic evaluations has been experienced between 1992 and 1996 are 'surgical' (from 10% to 17% of studies), 'diagnostic' (from 12% to 18% of studies) and 'care' (from 8% to 15% of studies). In addition to pharmaceuticals, the proportion of studies classified as 'prevention' also fell, from 17% to 9%. It is not possible to make a straightforward comparison with previous reviews of the literature, although Warner and Hutton (1980), using a narrower classification system of 'prevention', 'diagnosis' and 'treatment' also

observed an increase in the proportion of articles considering diagnosis and a decrease in the proportion of evaluations dealing with prevention, over time.

The relative, but not absolute, decline in pharmaceutical evaluations may be expected as interest in undertaking economic evaluations develops to encompass technologies other than pharmaceuticals. However, it has been noted that, for a long period, economic evaluation was 'a largely model based public sector funded activity' (Luce, 1995) and that the pharmaceutical industry entered into the field of economic analysis only in the late 1980s. It is therefore perhaps the case that concern with the evaluation of non-pharmaceutical technologies has now returned to its former dominance. Whatever the interpretation of this trend, the distinction between pharmaceuticals and other types of technology is of interest for a number of reasons. One of these relates to the sponsors of economic studies, given that it might be expected for the pharmaceutical industry to be the predominant sponsors of pharmaceutical evaluations, and given that doubt has been cast on the value of studies funded from this source (Freemantle and Maynard, 1994).

4.3 Sponsorship of studies

This issue was investigated by examining the sponsorship of all applied studies and those which involve pharmaceuticals, whether they be evaluations of pharmaceuticals alone or in comparison with another type of technology, with the results presented in table two. As a proportion of all economic evaluations for which a sponsor was stated, government and publicly funded bodies were the principal source of funds, sponsoring 36% of studies in 1996 compared with the pharmaceutical industry in second place with 32%. These organisations held the same positions with respect to the funding of studies in 1992 although the difference was greater (47% versus 38%). In comparison, pharmaceutical evaluations for 1996 show the positions reversed, with the pharmaceutical industry sponsoring 54% of studies for which a sponsor was identified and government and publicly funded bodies 34% (58% versus 38% of the small number of studies identified in 1992). It should be borne in mind that a sponsor was identified for only a proportion of studies; just under a third of all applied studies and of pharmaceutical evaluations had an identified sponsor in 1996, compared with around 40% in 1992. In addition, studies may be sponsored by more than one type of organisation. Nevertheless, there is an issue which is capable of further exploration concerning the methodological quality of studies funded by different organizations. Using the data discussed here, it would be possible to attempt to judge the methodological soundness of pharmaceutical industry-sponsored research relative to research sponsored by other organisations such as public bodies.

Another aspect of pharmaceutical industry-sponsored studies and pharmaceutical evaluations in general which is of interest when attempting to judge the quality of studies relates to the study design. Again, there are reasons for supposing that study design might differ as between evaluations of pharmaceuticals and other technologies. In particular, it might be thought that pharmaceutical evaluations would be more likely to exploit the randomized controlled trial, since this is the model for the establishment of the efficacy of pharmaceuticals. It would be of interest, therefore, to explore whether a decline in the proportion of studies considering pharmaceuticals has had an impact

Table 2 **Sponsors of Applied Studies - %ages**

Percentages of studies with declared sponsor	All evaluations		Pharmaceutical evaluations	
	1992	1996	1992	1996
Pharmaceutical industry	38%	32%	58%	54%
Govt./publicly funded policy making body	47%	36%	38%	34%
Charity	13%	17%	17%	12%
Research council/university	6%	29%	0%	24%
Public health care institution	4%	4%	4%	1%
Private health care institution	0%	2%	0%	1%
Non-pharmaceutical industry	4%	8%	0%	0%
Number of studies with declared sponsor	47	344	24	154

on the extent to which economic evaluations, in general, are conducted on the basis of randomized trials, particularly given concern with the methodological aspects of conducting economic evaluations alongside clinical trials.

4.4 Study types

It was possible to identify studies for which data on outcomes and on resource use had been collected as part of a single primary study. This was done by focusing on the data sources section of the review format, as illustrated in box two. Economic evaluations conducted alongside clinical trials were identified by selecting all those studies in which data sources for probabilities of main clinical events, quantities of resource use and outcomes are recorded as 'randomized clinical trial' with no other data source recorded. This is a particularly strict definition of what constitutes an economic evaluation alongside a clinical trial, and the conclusions drawn should be regarded as indicative rather than definitive. For example, those studies which have used a clinical trial for the purposes of assessing efficacy but have used observational data on resource use, such as in the case of multinational trials which rely on local out-of-trial estimates of costs, would be excluded, as would evaluations based on the results of randomized trials drawn from previously published sources, or those that used modelling and observational data to move from cost-efficacy to cost-effectiveness.

The results given in Table 3 show that, for all applied studies, the proportion falling into the category of economic evaluations alongside randomized trials has fallen by about half over the period, from 18% in 1992 to 10% in 1996. Thus, on the basis of the strict definition set out above, this category of study has not become more important. It has declined substantially in importance as a proportion of all economic evaluations. This trend applies almost equally to pharmaceutical evaluations and non-pharmaceutical evaluations. For pharmaceutical studies, the proportion accounted for by randomized trials, while being higher than for applied studies as a whole, has fallen from 23% to 11% between 1992 and 1996. It should be noted, however, that the absolute numbers of

Box 2

Probability of Main Clinical Events:

Randomised Clinical Trial: ☐	Systematic Review and/or Meta Analysis: ☐	Judgement: ☐
Observational Data: ☐	Other Literature Review: ☐	Modelling: ☐

Quantities of Resourcee Use:

Randomised Clinical Trial: ☐	Systematic Review and/or Meta Analysis: ☐	Judgement: ☐
Observational Data: ☐	Other Literature Review: ☐	Modelling: ☐

Prices or Costs of Resources:

Specific Estimates: ☐	Local Standard Costs: ☐	Judgement: ☐
National Publication: ☐	'Ad Hoc' Estimation: ☐	Local Standard Prices: ☐

Outcomes:

Randomised Clinical Trial: ☐	Systematic Review and/or Meta Analysis: ☐	Judgement: ☐
Observational Data: ☐	Other Literature Review: ☐	Modelling: ☐

Value of Outcomes:

Direct Valuation Within Study: ☐	Published Multi-attribute Utility Scale: ☐	Judgement: ☐
Valuation of Study Specific Scenarios: ☐	Previously Published Values: ☐	Willingness to Pay: ☐

Table 3 **Applied studies alongside Randomized Controlled Trials (RCTs)**

Percentages of all in group	1992	1993	1994	1995	1996
All evaluations alongside RCTs	18%	14%	13%	11%	10%
Pharmaceutical evaluations alongside RCTs	23%	16%	14%	17%	11%
All cost minimisation (CMA), cost effectiveness (CEA), cost utility (CUA), cost benefit analysis (CBA) alongside RCTs	16%	22%	15%	14%	10%
Pharmaceutical CMA, CEA, CUA, CBA alongside RCTs	23%	25%	14%	21%	11%
Nos. of applied studies	111	163	398	1038	1053
Nos. of pharmaceutical evaluations	60	86	186	407	482
Nos. of CMAs, CEAs, CUAs and CBAs	67	87	171	351	353
Nos. of pharmaceutical CMAs, CEAs, CUAs and CBAs	39	53	90	156	188

economic evaluations alongside randomized trials has increased substantially between these years, to five times their original level for all technologies and to four times for pharmaceutical evaluations.

Economic evaluations using only primary observational data were also identified, defined as where observational data was the only data source recorded for probability of main clinical events, resource use and outcomes. As a proportion of all applied studies, these observational studies increased from 24% to 39%. Associated with this has been an increase in the importance of cost consequences analyses, which provide data on costs and outcomes but do not synthesise the two, as a form of economic evaluation. Observational studies, defined in this way, accounted for over half of all cost consequences analyses in 1996 compared with around 30% in 1992. The increase in numbers of cost consequences analyses, based on observational data, would appear to have accounted for some of the decline in the proportion of economic evaluations conducted alongside clinical trials. However, when the analysis is limited to the four conventional types of economic evaluation of cost minimisation, cost effectiveness, cost utility and cost benefit analysis, a qualitatively, if not quantitatively, similar trend is shown for all applied studies. Thus, the proportion of these four types of evaluation that were conducted alongside randomized trials declined from 16% in 1992 (and over 20% in 1993) to 10% in 1996. The decline was more pronounced for pharmaceutical evaluations, with 11% of studies being conducted alongside clinical trials in 1996 compared with 23% in 1992. However, this series showed no clear trend, as the proportion went from 14% in 1994 to over 20% again in 1995. It can be concluded, however, that there has been no increase in the proportion of studies conducted alongside RCTs.

The collection of clinical and resource use data from different sources represents just one potential source of difficulty with making comparisons between studies, both within and between programme areas. In the latter case, the use of different outcome measures also makes comparisons difficult. In principle, the problem of different outcome measures can be overcome by the use of a common measure such as the quality adjusted life year (QALY) or the use of willingness to pay (WTP) in cost benefit analysis. However, in these studies, there may be variation in the way in which quality adjustments are applied or WTP is calculated. Given the small number of cost benefit analyses which have been conducted, the following section focuses on the different methodologies applied to the question of estimating utility measures in cost utility analysis.

4.5 Utility Measurement

By far the most common utility measure used is the quality adjusted life year (QALY). Only one instance of the Healthy Years Equivalent (HYE)

being used in an applied study was recorded during the period, compared with two instances in which the disability adjusted life year (DALY) was used. HEED records both the valuation method and the source of utility information. The method of valuation may be by asking patients within the study to rate their own quality of life (direct valuation), by presenting health state scenarios to individuals other than the patients themselves (valuation of study specific scenarios), by using a published multi-attribute utility scale, by the use of previously published values or by the use of judgement (of the authors or another group). More than one technique may be applicable to any one study. For example, more than one descriptor may be applied to a rating exercise (such as direct valuation using a published multi-attribute utility scale) or a variety of estimates may be derived using different approaches. Meanwhile, the source of utility data, which again can be classified under more than one category, can be from patients, from another source within the study (such as the authors) or from another study. Table 4 provides information on these two aspects of the data.

As far as the method of valuing outcomes is concerned, direct valuation within the study was used relatively infrequently, while previously published values and judgement tended to be the most frequently used methods. The distributions are not considered here for all the years 1992 to 1996 since the population sizes for 1992 to 1994 are relatively small. However, of the cost utility studies appearing in 1996, 43% used previously published values, 47% used judgement and only 13% used direct valuation. 22% and 29% used study specific scenarios and published multi-attribute utility scales, respectively. Although the results have not been broken down as in the study by Gerard (1992) to indicate what combinations of methods were used, they appear to confirm her earlier findings about the general lack of use of direct observation as opposed to the use of judgement. The results from HEED for the source of utility data also reflect the lack of direct valuation and the use of literature sources, as only 18% of cost utility studies in 1996 used patients as the source of utility data, compared with 45% which used another study. Around half of the studies used utility data from within the study but from sources other than patients, reflecting the widespread use of judgement to apply quality adjustments to life years.

Quality adjustments therefore appear to have been made in a relatively unsophisticated way for the purpose of estimating utilities in cost utility analysis, although economists have less to say about the merits of different approaches to applying a quality of life adjustment to life years than about other aspects of economic evaluation. Where there is an element of uncertainty about the appropriate method to use, economists will, however, always recommend the use of sensitivity analysis to investigate the impact of changes in method. The next section will, therefore, consider the use of sensitivity analysis, not just for cost utility studies but for all applied evaluations in HEED.

Table 4 **Valuation of Outcomes-Cost Utility Analyses (CUAs) (%ages)**

Method of valuation:	1995	1996
Direct valuation within study	18%	14%
Valuation of study specific scenarios	16%	22%
Published multi-attribute utility scale	18%	29%
Previously published values	39%	43%
Judgement	41%	47%
Source of data:		
Patients in study	18%	18%
Elsewhere in study	36%	51%
Incorporated from another study	41%	45%
Total number of CUAs	44	51

4.6 Sensitivity analysis

Table 5 indicates the proportions of studies which have investigated the impact of uncertainty through the use of sensitivity analysis from 1992 to 1996. Overall, the proportion of studies using sensitivity analysis has declined from around a third to about a quarter over the period. This is another area where the growth in cost consequences analyses has had an impact, as can be shown by comparing the use of sensitivity analysis for different types of economic evaluation. While over half of cost effectiveness analyses made allowance for uncertainty in 1996, and 85% of cost utility analyses, less than 10% of cost consequences analyses did so. It is perhaps unsurprising, and encouraging, that those studies which attempted to bring together information on costs and effects in a formal way (and would therefore satisfy a narrower definition of economic evaluation) should be more likely to allow for uncertainty by the use of sensitivity analysis. While it is not, in general, possible to identify from the database those variables which have been used to explore the impact on the results of uncertainty, it is possible to determine whether or not a range of estimates has been used for one variable in particular, the discount rate. This is the subject of the next section.

4.7 Discounting

The use of discounting, as with sensitivity analysis, can be taken as an indicator of the quality of an economic evaluation. Indeed, it might be argued that the discount rate should be one of the key parameters for which a range of values is included in the analysis since the discount rate is an assumed rather than an observed variable. Moreover, there is uncertainty about the appropriate rate to be used depending on the setting of the study and the viewpoint of the decision maker using the results of an economic evaluation. Hence, while the US Panel on the Use of Cost-Effectiveness in Health and Medicine recommend a rate of 3% in the reference case, a 'reasonable range' of values between 0% and 7% is recommended, with 5% being used in addition to 3% for at least the next ten years to maintain comparability with existing analyses (Gold et al., 1996). In the UK the public sector discount rate is 6%.

As an indication of the extent to which studies in HEED have used discounting, the results in Table 6 for cost minimisation analyses (CMA), cost effectiveness analyses (CEA), cost utility analyses (CUA) and cost benefit analyses (CBA) in 1996 show that a minority of studies discount either costs or benefits, with 23% discounting costs and 20% discounting benefits. It is not possible to identify the extent to which costs or benefits are left undiscounted because they are measured over a period of no more than one year, but it is expected that this would be the reason for not discounting in a number of studies which have discounted neither costs nor benefits. Another reason for not discounting benefits could be a decision on the part of the researchers that this is inappropriate, as argued by Parsonage and Neuberger (1992). One group of studies which could be examined to assess the extent to which this reasoning has been used would be those studies which include effectiveness (cost minimisation, consequences, effectiveness, utility or benefit analyses) and discount costs but not effects. In addition to noting that discounting has been used in a minority of studies, it is also notable that the discount rate was infrequently subject to sensitivity analysis. Of those CMA, CEA, CUA and CBA studies in 1996 which discounted costs, over 75% used a single discount rate (excluding some studies which presented undiscounted results in addition to using a positive discount rate). For those studies which discounted effects, about 70% used a single rate. Including studies which subjected the discount rate to sensitivity analysis, the most common rate was 5%, with around 70% of studies which discounted costs using this rate and over 75% of those which discounted effects.

Table 5 **Sensitivity analysis - applied studies (%)**

	1992	1993	1994	1995	1996
Sensitivity tested	32%	33%	24%	18%	23%
Quantitatively reported	31%	33%	22%	16%	22%
No. of applied studies	111	163	398	1038	1053

Table 6 **Discounting - %age of applied studies**

	1992	1993	1994	1995	1996
Costs discounted - cost minimisation (CMA), cost effectiveness (CEA), cost utility (CUA) and cost benefit analysis (CBA)	24%	30%	20%	16%	23%
Benefits discounted - CMA, CEA, CUA, CBA	19%	28%	16%	13%	20%
Numbers of CMA. CEA, CUA and CBA	67	87	171	351	353

5. QUALITY OF STUDIES

While no formal attempt has been made in this paper to evaluate the quality of studies in HEED, it is possible to draw out some conclusions from the information presented. What indicators there are suggest that the quality of studies may not have improved at the same rate as the rate of increase in the numbers of studies over the period 1992 to 1996. For example, the study design with the greatest internal validity, the economic evaluation conducted alongside a randomized controlled trial, has on the strict definition used in this paper become a less important means of conducting economic appraisals as a proportion of applied studies. Of course it may be that a larger proportion of studies are based on trial data but combine it with other information, in which case they would fall outside of our definition. The trends with regard to the use of sensitivity analysis and discounting also do not, however, provide grounds for believing that the quality of studies is improving.

A more detailed analysis of a subset of studies is required to draw firmer conclusions about the quality of studies, perhaps along the lines of the work carried out by previous researchers. Box three provides a summary of a number of studies, all published since 1990, which have attempted to provide a detailed analysis of the quality of a selected group of studies typically using a checklist approach. Trakas et al. (1997) and Taddis et al. (1994) have been excluded because they evaluated abstracts only. The study by Zarnke et al. (1997) was also excluded because a checklist was not used, although the overall comprehensiveness of studies was rated. Seven of the eight studies included concerned themselves with the overall quality of the chosen sample of studies while the other (Briggs and Sculpher, 1995) focused specifically on the way in which uncertainty was handled. Since differences in methodology may influence any attempt to make comparisons between these exercises as much as between economic evaluations themselves, the methods as well as the results of the various studies have been summarised.

5.1 Previous reviews of study quality

It is difficult to make overall judgements on the quality of economic evaluations on the basis of the studies presented in box three because differences in methodology limit the extent to which comparisons can be made between studies. Each study has applied a different checklist to a subset of economic evaluations selected on the basis of varying criteria, whether they be according to journal type, type of economic evaluation (Gerard, 1992), or according to a statement of intent by the author(s), rather than an objective assessment of whether an economic evaluation has been conducted (Udvarhelyi, 1992).

Some studies provide an overall assessment while others provide an assessment on individual items. Among those studies which made an overall assessment, Bradley et al. (1995) found a 73% adequacy rate across all checklist items, Gerard (1992) found 63% of 'worthwhile' studies to be technically average or above average, and Adams et al. (1992) found a mean completeness score of 0.52 (scaled from 0 to 1). These results may suggest that 50% or more of studies reach a reasonable level of technical quality. There are, however, discrepancies between studies on particular aspects of economic evaluation. For example, Briggs and Sculpher (1995) found that 77% of their studies had dealt with uncertainty, albeit inadequately in the majority of cases, whereas Lee and Sanchez (1991) found that less than 10% of their studies undertook a sensitivity analysis while, in the studies examined by Adams et al. (1992), only 16% subjected costs to sensitivity analysis, with 8% doing so for benefits. On the other hand, 30% of articles examined by Udvarhelyi et al. (1992) are reported as having used sensitivity analysis, and 86% of Gerard's (1992) sample carried out a sensitivity analysis, although this was more likely to be limited than extensive.

Box 3

1. Ganiats and Wong (1991)

Method: MEDLINE search using MeSH heading 'cost benefit analysis' and manual search to identify articles published in six medical journals between January 1982 and November 1987. 47 articles were compared on seven criteria, covering the comprehensiveness of the cost assessment, measurement and valuation of outcomes, use of discounting and sensitivity analysis and appropriateness of conclusions. 'Benefit-cost' analyses (BCAs), expressing the results as a net resource cost (none valued health outcomes in monetary terms), and cost-effectiveness analyses (CEAs), deriving a ratio of resource cost to health outcome, were considered.

Results: Ratings on the various criteria depended on the proportion of articles satisfying the criteria, with 'high' for 70% or more, 'medium' for 50% to 69%, 'low' for 10%-49% and '0' for less than 10%. While all studies included the immediate direct costs, with both BCAs and CEAs scoring 'high' on this criterion, over one third neglected the costs of further workup or treatment, both types of study scoring 'medium'. None of the BCAs and only 15% of the CEAs calculated health care costs in added years of life, giving '0' and 'low' marks respectively. A 'high' mark was achieved for the proportion of CBAs and CEAs discounting future dollar costs and benefits but two studies discounted costs but not health outcomes and a 'low' mark was given to each type of study when the presentation of undiscounted results was considered. 35% of BCAs and 70% of CEAs subjected key variables to sensitivity analysis, giving 'low' and 'high' marks respectively, and fewer the discount rate, BCAs being marked 'low' and CEAs 'medium'. 80% of BCAs ('high') but only 30% of CEAs ('low') had appropriate conclusions.

2. Lee and Sanchez (1991)

Methods: Searches of MEDLINE, Health Planning and Administration and International Pharmaceutical Abstracts to identify publications dealing with economic issues in six pharmacy journals between January 1985 and December 1990. An 10-point checklist adapted from Drummond et al. (1987) was applied to 65 studies asserting the cost-effectiveness of a drug or pharmaceutical services intervention.

Results: Only three of the ten criteria were satisfied by more than half the studies. These were 'competing alternatives described comprehensively' (71%), 'costs and consequences measured accurately in appropriate physical units' (85%) and 'costs and consequences valued credibly' (54%). Only 31% of articles identified all relevant costs and consequences for each alternative. Only one study (2%) performed an incremental analysis, and only 6% of studies undertook a sensitivity analysis. It was concluded that basic methodological aspects of economic evaluation were frequently overlooked.

3. Adams et al. (1992)

Methods: Search of MEDLINE from January 1966 to June 1988 to identify randomized trials which considered costs or economic analysis. 51 of the 121 articles fulfilling the criteria were assessed. A checklist for completeness, again based on that presented by Drummond et al. (1987), was applied. For those aspects of the checklist seeking to ascertain whether a study had conducted the analysis appropriately, e.g. was an appropriate sensitivity analysis conducted correctly, a rating of 0,1 or 2 was given, corresponding, in this example, to 'complete and correct analysis', 'partial sensitivity analysis' and 'no sensitivity analysis'. The completeness score was expressed on a 0-1 scale as the proportion of total possible points gained across the checklist.

Results: The mean score for completeness of the economic evaluation on the 0-1 scale was 0.52, with a range of 0.32 to 0.94. Those studies which were designed to incorporate economic analysis prospectively and those which included the appropriate costs and benefits received higher score. An explicit statement of the perspective or viewpoint was made in only one quarter of studies, although it could be deduced in 43% of studies in which it was not explicitly specified. 38 studies included a prospective economic evaluation in the original protocol, of which 26 did not include appropriate measures of costs or consequences. Neither was this the case in those 13 studies which appeared to be retrospective. 84% of studies did not subject costs to sensitivity analysis, and 92% did not do so for benefits.

4. Gerard (1992)

Methods: Searches of a variety of published and unpublished literature sources and contact with researchers working on cost-utility analysis to identify such studies, that is, studies reporting a cost per QALY ratio and reported in the period 1980 to mid 1991. A set of 40 characteristics/criteria developed with the aid of published checklists and expert opinion was used to evaluate each of 51 studies, with an attempt made to obtain information from the relevant researchers if necessary. Technical criteria covering the requirements of a comprehensive cost-utility analysis, comprised 23 items; in addition, an overall assessment of quality and value of studies was made.

Results: Of the 51 studies considered, 45 included only direct costs. 29 studies reported the use of marginal costs, and another six the use of marginal and average costs. More than two-thirds covered the relevant cost areas, two studies did not and, in 14, no

judgement could be made. 14 studies were unclear as to the process of measuring costs. For 11 studies, the sources of cost data were not identified, and for a further nine, the information was partial. In 11 studies, the price base was not made explicit. The quality adjustment was carried out using general health outcome measures in 33 studies, using disease specific measures in 15, while for three it was unclear. Clear or partial descriptions of how life years gained were estimated were provided in 38 studies; a combination of 'estimates' (from reliable sources such as published literature and informed expert opinion) and 'guesses' (by the researcher(s) or over expert opinion) were used in 27 cases, with 'estimates' alone used in four and 'guesses' alone in 2. Quality of life effects were estimated in this way for 17 studies, with guesses used in another four and direct observation used in only 16 studies. The community was used as the most common source of QALY valuations (21 studies); in 14 it was not possible to tell and in 4 the researchers' values were applied. In 36 studies, both costs and outcomes were discounted, while in three only costs were discounted. In seven in which discounting was appropriate, none was undertaken with no reason given or it was unclear whether it had been done. Sensitivity analysis was judged to be limited in 21 studies, based on the justification given by the authors for the values of key variables. The study design (source of effectiveness data) was also unclear in 21 cases. Overall, 46 studies were considered worth undertaking, with 16 technically average, 13 above and 17 below average. Four were not worth undertaking, although two were of good technical quality, and no judgement could be made on remaining study.

5. Udvarhelyi et al. (1992)

Methods: MEDLINE search to identify articles concerned with a focus on cost-effectiveness or cost-benefit issues and published between 1978 and 1980 or 1985 and 1987 in a group of general medical, general surgical or medical sub-specialty journals. The focus of the article was based on 'cost benefit' or 'cost effective(ness)' in the title or a statement in the article that the study intended to compare costs and effects. Six basic principles were applied to the resulting 77 articles.

Results: 4% of articles satisfied all six principles of analysis, with another 19% satisfying five principles. 18% satisfied the first principle by making an explicit statement of the perspective used. For principle two, 83% of studies listed the benefits under consideration but, for 14 studies, these could not be ascertained. On principle three dealing with the statement of costs being considered, 3 (4%) provided no assessment of costs (despite a statement of the intention to consider costs), the same number as included all components of cost. 9% were too ambiguous to determine whether programme costs had been included, and 30% took no account of costs of side effects or morbidity or cost savings. Of 29 articles covering a time period of more than a year, only 14 used discounting appropriately (principle four) and only 30% of articles used sensitivity analysis. 42% of articles satisfied principle six by reporting a cost-effectiveness or cost benefit ratio. This literature showed poor compliance with some basic principles of economic evaluation.

6. Bradley et al. (1995)

Methods: Search of MEDLINE, EMBASE and International Pharmacy Abstracts to identify articles published between January 1989 and December 1993 in three health economics, three medical and three pharmacy journals. Articles were selected if they considered costs and outcomes and could be classified as cost-effectiveness, cost-utility, cost-benefit or cost-minimization analyses. Editorials, commentaries and reviews were excluded. The 12-item checklist of Sacristan et al. (1993) was applied to a total of 90 articles. This checklist deals with, for example, the definition of study aim, the analysis of alternatives and of perspective, the measurement of costs and of benefits, analysis of results and discussion of the assumptions and limitations of the study. On each item, a score of 4 (correct), 3 (acceptable), 2 (doubtful), 1 (not reported) or 0 (incorrect) was allocated. The sum of the individual scores was divided by the number of applicable items to derive an overall score. A 13th item (overall impression) was used as a validity check.

Results: 73% of responses across all 12 items were rated adequate, scoring either three or four. Mean overall scores increased from 2.5 in 1989 to 3.2 in 1993, with all but one item (discussion of ethical problems) being rated adequate (correct or acceptable) in over 75% of studies in the latter year. Thus, all studies in 1993 were rated adequate according to definition of study aim, appropriateness of sample selection, suitability of evaluation if within a clinical trial and justification/generalization of conclusions. The adequacy and relevance of benefit measurement and the suitability of cost measurement are also rated highly, receiving an adequate rating in 76% and 94% of studies in 1993. This finding is perhaps surprising in light of previous findings given that cost measurement includes a consideration of adjusting for future costs. Similarly, discussion of assumptions and study limitations has a surprisingly high adequacy rating of 88% given that this item includes the performance of sensitivity analysis. However, a rating of 'adequate' blurs the distinction between 'acceptable' and 'correct'; item four (analysis of perspective) was rated as adequate in 67% of studies but as correct in only 16% because of the requirement to state the perspective explicitly to be 'correct'.

7. Briggs and Sculpher (1995)

Methods: Search of MEDLINE for 1992 for articles with economic evaluation terms in the title (excluding editorials, letters or reviews), plus a manual search of the Wellcome Economic Evaluation Bibliography (excluding methods papers) and of Health Economics and PharmacoEconomics. The requirement for empirical evaluations of health care interventions providing a complete analysis of costs and outcomes gave 93 studies in total. A set of questions to be asked of each study was developed dealing specifically uncertainty, four aspects of which were examined. The synthesis of study results and overall quality of the studies were also assessed.

Results: The overall quality assessment indicated that, of the 93 studies assessed, 13 made a good attempt to allow for uncertainty. This was a subjective judgement, indicating that the study was deemed to have made a detailed attempt to assess the impact of uncertainty on the study results. 23 studies were rated adequate, that is, the sensitivity analysis had dealt with the major areas of uncertainty while, in 35 studies, uncertainty was inadequately dealt with since major areas of uncertainty were not considered or it was not clear to what extent sensitivity analysis had been carried out. The other 21 studies failed to take any account of uncertainty. Thus, 77% of studies made some allowance for uncertainty. One particularly weak aspect was that 51 studies did not explore the impact of uncertainty with regard to unit costs; in 23 studies the source of this information was not made clear. In terms of generalizability, only 17 of 63 studies which recognised the possible relevance of the results beyond their own context failed to conduct sensitivity analysis; however, in only 10 studies was this comprehensively performed. Most studies in which uncertainty related to the inclusion or exclusion of indirect costs failed to investigate the impact of altering the approach. Only three studies assessed the impact of health care costs in added years of life using sensitivity analysis. 42 studies undertook a one-way sensitivity analysis, 15 a multi-way sensitivity analysis and only five an extreme scenario analysis. Only 15 studies reported the sources from which the ranges of variables employed had been taken. It is concluded that, in practice, sensitivity analysis is less comprehensive than the ideal level recommended in the methodological literature.

8. Blackmore and Magid (1997)

Methods: Medline search of 21 major peer-reviewed diagnostic radiology journals, for original research articles with 'cost-effective', 'cost-effectiveness', 'cost-benefit' or 'benefit-cost' in the title or abstract or subject heading 'cost-benefit analysis' and in which economic evaluation was a primary focus. The 44 studies identified were assessed according to six major and four minor methodologic principles adapted from Udvarhelyi et al. (1992).

Results: Five of the 44 articles satisfied all six major criteria, of which three also satisfied the minor criteria. Studies satisfied a median of three major criteria and one minor criterion. Of the major criteria, most often satisfied were an explicit statement of the diagnostic or therapeutic options considered and a clear statement of the estimated cost (43 studies each or 98%) and an explicit statement of outcomes (35 studies or 80%). A statement of the source of cost data, a minor criterion, was also frequently satisfied, in this case by 31 studies (70%). 18% of studies performed a sensitivity analysis, 14% provided an explicit statement of the study's perspective and 11% used discounting.

It may or may not be possible, therefore, to extrapolate the conclusions of the studies summarised in Box 3 to economic evaluations generally. Moreover, the vintage of the evaluations included in these studies may preclude any conclusions being drawn about the current state of the art. Given that no evaluations beyond 1993 have been assessed in this group of studies, and that most are from well before this date, it would be appropriate to conduct a quality assessment exercise on a more recent sample of evaluations to investigate whether or not the quality of studies has changed over time. The following section will consider some of the problems of conducting such an exercise.

5.2 Problems of quality assessment

One of the difficulties of judging the quality of economic evaluations is the lack of an objective measure of quality. Those studies which have applied a checklist to a group of studies have used a variety of different approaches to assessing quality and have typically assessed quality on the basis of whether or not a study has included elements of economic evaluation which the researchers consider important. It is, therefore, possible to compare the quality of studies according to the proportion of items in the checklist that are included in the analysis. However, this assumes that each component is equally important, when some may be more important than others. Some factors which are traditionally included in a checklist may have more to do with the detail with which the results are reported rather than the quality of the underlying study. For example, the use of a

particular discount rate may be relatively unimportant to a consumer of an economic evaluation if, given the raw data, the impact of using any discount rate selected by the user can be evaluated. Reinhardt (1997) emphasizes the importance of providing access to the raw data of studies, seldom emphasized by traditional approaches to methodological standards.

Perhaps most important would be the quality of the underlying data used in a study. The US Cost-Effectiveness Panel identified a range of different study types and recommended that those sources with the least potential for bias should be used, ranging from a randomized trial, with the least bias to authors' judgement, with the most potential for bias. However, while randomized trials may have the greatest internal validity, their external validity may be brought into question. In particular, it has been pointed out that resource use as measured in randomized trials may vary from that relevant to usual clinical practice (Drummond and Davies, 1991). A possible alternative, to use a modelling study based on a variety of sources, is not itself without drawbacks. However, there is uncertainty in the economic evaluation literature about the relative merits of trial-based and modelling-based studies (Sheldon, 1996) and this is reflected in the checklists and guidelines on economic evaluation (Towse, 1997). This is perhaps, in part, because the methodology of modelling studies is less well defined than that of randomized trials.

For example, the guidance produced by the US Cost-Effectiveness Panel suggests that models should be subject to validation, but does not indicate what types of models have demonstrated validity nor does it suggest a standard approach to validity testing. In addition, while the methodology for comparing and combining the results of randomized trials in meta-analysis (and attempts are underway to undertake a similar analysis for economic evaluations), there is no corresponding methodology for combining the results of modelling studies. It would be helpful, for those attempting to make comparisons between studies, to have firmer guidance taking account of the clinical and economic data sources on which economic evaluations are based. The following conclusions section will consider ways in which the use of a source of evaluations such as HEED could help to advance an understanding of the quality of economic evaluations.

6. CONCLUSIONS AND FURTHER RESEARCH

The first conclusion to draw from this review of the HEED database is that the annual numbers of economic evaluations being conducted has continued to expand beyond 1992, the point beyond which relatively little information has been available to date. The information presented here suggests that the annual number of articles may have increased substantially after 1994, although this will be in part an artefact of improved coverage of the database over time and not simply a reflection of the underlying growth in economic studies. In addition to an increase in the quantity of literature covered, an improvement in the quality of the reviews entered on the database may have influenced some of the other trends identified, for example, the distribution of studies by ICD-9 and ATC chapter. With this caveat in mind, there are trends which can be identified with a good deal of confidence. One is the marked shift over the period towards applied studies (original attempts to bring together on costs and outcomes, on the cost of a disease or the relative costs of two or more approaches to treatment for a given condition).

It is likely that the increased emphasis on original studies has occurred due to factors that influence demand from funding organisations and editors of journals, perhaps in response to conclusions of early reviews that more, better quality, studies are required and factors that influence the willingness of researchers to undertake economic evaluations. An increasing appreciation of the importance of economic considerations on the part of those who commission research has perhaps encouraged researchers, in particular clinical researchers, to use this type of study as a vehicle for publishing original clinical research, with the assessment of costs being regarded as of secondary importance. The costs of conducting studies should also favour observational studies, particularly when based on retrospective review, compared with prospectively conducted randomized trials.

Another distinct feature of the data is that the increase in applied studies has come about predominantly through studies which have considered costs and outcomes separately without attempting a synthesis between the two (cost consequences analysis). This can be seen as an important development because, while these studies provide a large proportion of the information available for decision makers in health care concerned with using data on costs and outcomes, they may be of limited use. For example, it is difficult to make comparisons, even

within a single therapeutic area, between studies which do not yield an incremental (or even average) cost-effectiveness ratio. Even for a particular condition and procedure, studies may use multiple outcome measures which cannot be compared from one study to another. Moreover, where it has not been the intention from the outset to synthesise information on costs and outcomes, costs may be included as an afterthought and not assessed as rigorously as for other types of study.

Another factor limiting the usefulness of some cost consequences studies is that they do not all satisfy both requirements for a full economic evaluation of being a 'comparative analysis of alternative courses of action in terms of both their costs and consequences' (Drummond et al., 1997). While all studies evaluate both the costs and consequences of an intervention, it is not always the case that a comparator has been explicitly included in the analysis. Rather, some studies are concerned with the costs of treatment and the health state before and after treatment for a single intervention among a cohort of patients with a given condition, and would therefore fall into the category of cost-outcome descriptions. Unfortunately, it is not straightforward to identify which studies have employed an explicit comparator and which have not. However, it may be hypothesised that cost-outcome descriptions, since they are likely to be of relatively little interest to an economist, would be conducted on the grounds of clinical interest rather than to address an economic question. Indeed, one possible explanation for the observed increase in cost consequences analyses, which have tended to drive the increased emphasis on applied studies overall, is an increase in the level of interest in economics on the part of clinicians.

Since the increase in cost consequences analyses has occurred alongside a decline in the proportion of studies being conducted alongside randomized controlled trials (RCTs), (albeit on a strict definition) there is a question as to whether it might then be concluded that the overall quality of economic studies has declined. However, while it is generally agreed that the RCT provides estimates of costs and effects with the minimum degree of bias, economic evaluations alongside clinical trials do not provide all the answers (OHE, 1997). In future, therefore, economists may increasingly find it to their advantage to exploit the large amount of observational data available to supplement data gathered from trials. A useful research question for the future would be to assess not only whether trial-based evaluations continue to lose ground relative to other types of study but also how useful different types of study design are for providing good quality information.

On the subject of quality assessment, the large amount of data which exists on HEED could be used to assess the characteristics and quality of studies according to type of sponsor in order, for example, to identify whether or not studies sponsored by pharmaceutical companies comply with established guidelines for economic evaluation more or less than studies funded by other organisations. In the UK, the practice of economic evaluations funded by the pharmaceutical industry could be compared with the principles embodied by the joint guidelines of the Department of Health and the Association of the British Pharmaceutical Industry (Joint Government/Pharmaceutical Industry Working Party, 1994). It would also be of interest to compare the quality of studies based on modelling approaches with studies of the same interventions and disease areas based on RCTs to attempt to shed light on this important area of the methodological debate. Studies performed on common interventions and in common disease areas could be also identified for the purposes of reviewing the evidence in a particular therapeutic area, including pooling results between studies. Finally, revisiting this overview periodically would help to give an insight into the volume and the characteristics of studies over time and into whether the trends identified here can be expected to continue.

REFERENCES

Adams M E, McCall N T, Gray D T, Orza M J, Chalmers T C (1992). Economic analysis in randomized control trials. Medical Care 30(3): 231-243.

Backhouse M E, Backhouse R J, Edey S A (1992). Economic evaluation bibliography. Health Economics 1(Supplement): 1-236.

Blackmore C, Magid D J (1997). Methodologic evaluation of the radiology cost-effectiveness literature. Radiology 203: 87-91.

Bowie C, Beck S, Bevan G, Raftery J, Silverton F, Stevens A (1997). Estimating the burden of disease in an English region. Journal of Public Health Medicine 19(1): 87-92.

Bradley C A, Iskedjian M, Lanctot K L, Mittmann N, Simone C, St Pierre E, Miller E, Blatman B, Chabursky B, Einarson T R (1995). Quality assessment of economic evaluations in selected pharmacy, medical, and health economics journals. The Annals of Pharmacotherapy 29: 681-689.

Briggs A, Sculpher M (1995). Sensitivity analysis in economic evaluation: a review of published studies. Health Economics 4: 355-371.

Department of Health (1994). Register of cost-effectiveness studies. London: Department of Health.

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

Drummond M F, O'Brien B, Stoddart G L, Torrance G W (1997). Methods for the economic evaluation of health care programmes 2nd edition. New York: Oxford University Press.

Elixhauser A, Luce B R, Taylor W R, Reblando J (1993). Health care CBA/CEA: an update on the growth and composition of the literature. Medical Care 31(7, Supplement): JS1-JS11.

Freemantle N, Maynard A (1994). Something rotten in the state of clinical and economic evaluations? Health Economics 3: 63-67.

Ganiats T G, Wong A F (1991). Evaluation of cost-effectiveness research: a survey of recent publications. Family Medicine 23: 457-462.

Gerard K (1992). Cost-utility in practice: a policy-maker's guide to the state of the art. Health Policy 21: 249-279.

Gold M R, Siegel J E, Russell L B, Weinstein M C (1996) (eds). Cost-effectiveness in health and medicine. New York: Oxford University Press.

Joint Government/Pharmaceutical Industry Working Party (1994). UK guidance on good practice in the conduct of economic evaluations of medicines. British Journal of Medical Economics 7:63-64.

Koopmanschap M A, van Roijen L, Bonneux L, Bonsel G J, Rutten F F H, van der Maas PJ, The Technology Assessment Methods Project Team (1994). Cost of diseases in international perspective. European Journal of Public Health 4(4): 258-284.

Lee J T, Sanchez L A (1991). Interpretation of 'cost-effective' and soundness of economic evaluations in the pharmacy literature. American Journal of Hopsital Pharmacy 48: 2622-2627.

Luce B R (1995). Policy implications of modeling the cost-effectiveness of health care technologies. Drug Information Journal 29:1469-1475.

Mason J, and Drummond M (1995). The DH register of cost-effectiveness studies: content and quality. Health Trends 27(2): 50-56.

NHS Centre for Reviews and Dissemination (1996). Making cost-effectiveness information accessible: the NHS economic evaluation database project: CRD guidance for reporting ctitical summaries of economic evaluations. CRD Report 6. York: NHS Centre for Reviews and Dissemination.

Parsonage M, Neuberger H (1992). Discounting and health benefits. Health Economics 1: 71-76.

Reinhardt U E (1997). Making economic evaluations respectable. Social Science and Medicine 45(4): 555-562.

SCRIP (1997). World drug sales climb 7% in 1st qtr. SCRIP 2241: 16.

Sheldon T A (1996). Problems of using modelling in the economic evaluation of health care. Health Economics 5: 1-11.

Taddio A, Pain T, Fassos F, Boon H, Ilersich A L, Einarson T R (1994). Quality of nonstructured and structured abstracts of original research articles in the British Medical Journal, the Canadian Medical Association Journal, and the Journal of the American Medical Association. Canadian Medical Association Journal 150: 1611-1615.

Towse A (1997). Guidelines for the economic evaluation of pharmaceuticals: can the UK learn from Australia and Canada? London: Office of Health Economics.

Trakas K, Addis A, Kruk D, Buczek Y, Iskedjian M, Einarson T R (1997). Quality assessment of pharmacoeconomic abstracts of original research articles in selected journals. The Annals of Pharmacotherapy 31: 423-428.

Udvarhelyi S, Colditz, Rai A, Epstein A M (1992). Cost-effectiveness and cost-benefit analyses in the medical literature: are the methods being used correctly? Annals of Internal Medicine 116: 238-244.

Warner K E, Hutton R C (1980). Cost-benefit and cost-effectiveness analysis in health care. Medical Care 18(11): 1069-1084.

Zarnke K B, Levine M A H, O'Brien B J (1997). Cost-benefit analyses in the health-care literature: don't judge a study by its label. Journal of Clinical Epidemiology 50(7): 813-822.

Clive Pritchard is a Research Associate at the Office of Health Economics, London, with specific responsibility for the economic evaluations that appear on the HEED database.

The data used in this Briefing were drawn from the February 1998 release of the HEED database which is produced monthly on CD-ROM. Further information is available from:

Gerry Crosbie or Clive Pritchard at

HEED, 12 Whitehall, London SW1A 2DY.

Tel: 0171 930 9203 ext. 1458/1474

Fax: 0171 747 1419

E-mail: database@abpi.org.uk

Office of Health Economics

The Office of Health Economics was founded in 1962. Its terms of reference are to:

- commission and undertake research on the economics of health and health care;
- collect and analyse health and health care data from the UK and other countries;
- disseminate the results of this work and stimulate discussion of them and their policy implications.

The Office of Health Economics is supported by an annual grant from the Association of the British Pharmaceutical Industry and by sales of publications, and welcomes financial support from other bodies interested in its work.

Recent OHE publications

Managed Care - a Model for the UK? by Dr Peter West, 1998 (price £10.00)

Competition and Contestability between Acute Hospitals by Jon Sussex, 1998 (price £10.00)

Compendium of Health Statistics (10th Edition) by Sean McGuigan, 1997 [price £195.00 (£60.00 to public sector organisations) plus P&P for non-UK].

Theory and Evidence on Cost Sharing in Health Care: an Economic Perspective by Martin Chalkley and Ray Robinson, 1997 (price £10.00).

Prescribing, Budgets and Fundholding in General Practice by Darrin Baines, Keith Tolley and David Whynes, 1997 (price £7.50).

Drivers of the Growth in Medicines Expenditure by Nick Marchant, 1997 (price £10.00).

NHS Waiting Lists: Towards the Elusive Solution by Rhiannon Tudor Edwards, 1997 (price £7.50).

Regulating the Prices Paid for NHS Medicines: Lessons from Utility Regulation by Martin Cave and Adrian Towse, 1997 (price £5.00).

Guidelines for the Economic Evaluation of Pharmaceuticals: Can the UK learn from Australia and Canada? Edited by Adrian Towse, 1997 (price £10.00).

Health Economics: An Introduction to Economic Evaluation by Gisela Kobelt, 1996 (price £2.00).

A full list of OHE publications is available from the address below.

OHE publications are available from the Office of Health Economics, 12 Whitehall, London SW1A 2DY. To order, send a cheque or telephone (0171 930 9203) or fax (0171 747 1419) with credit card details.