Health Technology Assessment 2004; Vol. 8: No. 49

Generalisability in economic evaluation studies in healthcare: a review and case studies

MJ Sculpher, FS Pang, A Manca,
MF Drummond, S Golder, H Urdahl,
LM Davies and A Eastwood

December 2004

Health Technology Assessment
NHS R&D HTA Programme

How to obtain copies of this and other HTA Programme reports.
An electronic version of this publication, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (http://www.hta.ac.uk). A fully searchable CD-ROM is also available (see below).

Printed copies of HTA monographs cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our Despatch Agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per monograph and for the rest of the world £3 per monograph.

You can order HTA monographs from our Despatch Agents:

– fax (with **credit card** or **official purchase order**)
– post (with **credit card** or **official purchase order** or **cheque**)
– phone during office hours (**credit card** only).

Additionally the HTA website allows you **either** to pay securely by credit card **or** to print out your order and then post or fax it.

Contact details are as follows:

HTA Despatch
c/o Direct Mail Works Ltd
4 Oakwood Business Centre
Downley, HAVANT PO9 2NP, UK

Email: orders@hta.ac.uk
Tel: 02392 492 000
Fax: 02392 478 555
Fax from outside the UK: +44 2392 478 555

NHS libraries can subscribe free of charge. Public libraries can subscribe at a very reduced cost of £100 for each volume (normally comprising 30–40 titles). The commercial subscription rate is £300 per volume. Please see our website for details. Subscriptions can only be purchased for the current or forthcoming volume.

Payment methods

Paying by cheque
If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *Direct Mail Works Ltd* and drawn on a bank with a UK address.

Paying by credit card
The following cards are accepted by phone, fax, post or via the website ordering pages: Delta, Eurocard, Mastercard, Solo, Switch and Visa. We advise against sending credit card details in a plain email.

Paying by official purchase order
You can post or fax these, but they must be from public bodies (i.e. NHS or universities) within the UK. We cannot at present accept purchase orders from commercial companies or from outside the UK.

How do I get a copy of *HTA on CD*?

Please use the form on the HTA website (www.hta.ac.uk/htacd.htm). Or contact Direct Mail Works (see contact details above) by email, post, fax or phone. *HTA on CD* is currently free of charge worldwide.

The website also provides information about the HTA Programme and lists the membership of the various committees.

Generalisability in economic evaluation studies in healthcare: a review and case studies

MJ Sculpher,[1*] FS Pang,[1] A Manca,[1]
MF Drummond,[1] S Golder,[2] H Urdahl,[1]
LM Davies[3] and A Eastwood[2]

[1] Centre for Health Economics, University of York, UK
[2] Centre for Reviews and Dissemination, University of York, UK
[3] Manchester Medical School, Manchester University, UK

* Corresponding author

Declared competing interests of authors: none

Published December 2004

This report should be referenced as follows:

Sculpher MJ, Pang FS, Manca A, Drummond MF, Golder S, Urdahl H, *et al*. Generalisability in economic evaluation studies in healthcare: a review and case studies. *Health Technol Assess* 2004;**8**(49).

Health Technology Assessment is indexed in *Index Medicus*/MEDLINE and *Excerpta Medica*/EMBASE.

NHS R&D HTA Programme

The research findings from the NHS R&D Health Technology Assessment (HTA) Programme directly influence key decision-making bodies such as the National Institute for Clinical Excellence (NICE) and the National Screening Committee (NSC) who rely on HTA outputs to help raise standards of care. HTA findings also help to improve the quality of the service in the NHS indirectly in that they form a key component of the 'National Knowledge Service' that is being developed to improve the evidence of clinical practice throughout the NHS.

The HTA Programme was set up in 1993. Its role is to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies is produced in the most efficient way for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined to include all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care, rather than settings of care.

The HTA programme commissions research only on topics where it has identified key gaps in the evidence needed by the NHS. Suggestions for topics are actively sought from people working in the NHS, the public, consumer groups and professional bodies such as Royal Colleges and NHS Trusts.

Research suggestions are carefully considered by panels of independent experts (including consumers) whose advice results in a ranked list of recommended research priorities. The HTA Programme then commissions the research team best suited to undertake the work, in the manner most appropriate to find the relevant answers. Some projects may take only months, others need several years to answer the research questions adequately. They may involve synthesising existing evidence or designing a trial to produce new evidence where none currently exists.

Additionally, through its Technology Assessment Report (TAR) call-off contract, the HTA Programme is able to commission bespoke reports, principally for NICE, but also for other policy customers, such as a National Clinical Director. TARs bring together evidence on key aspects of the use of specific technologies and usually have to be completed within a limited time period.

Criteria for inclusion in the HTA monograph series
Reports are published in the HTA monograph series if (1) they have resulted from work commissioned for the HTA Programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

The research reported in this monograph was commissioned by the HTA Programme as project number 98/22/05. As funder, by devising a commissioning brief, the HTA Programme specified the research question and study design. The authors have been wholly responsible for all data collection, analysis and interpretation and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the referees for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

The views expressed in this publication are those of the authors and not necessarily those of the HTA Programme or the Department of Health.

ISSN 1366-5278

Published by Gray Publishing, Tunbridge Wells, Kent, on behalf of NCCHTA.
Printed on acid-free paper in the UK by St Edmundsbury Press Ltd, Bury St Edmunds, Suffolk.

M

Abstract

Generalisability in economic evaluation studies in healthcare: a review and case studies

MJ Sculpher,[1*] FS Pang,[1] A Manca,[1] MF Drummond,[1] S Golder,[2] H Urdahl,[1] LM Davies[3] and A Eastwood[2]

[1] Centre for Health Economics, University of York, UK
[2] Centre for Reviews and Dissemination, University of York, UK
[3] Manchester Medical School, Manchester University, UK
* Corresponding author

Objectives: To review, and to develop further, the methods used to assess and to increase the generalisability of economic evaluation studies.
Data sources: Electronic databases.
Review methods: Methodological studies relating to economic evaluation in healthcare were searched. This included electronic searches of a range of databases, including PREMEDLINE, MEDLINE, EMBASE and EconLit, and manual searches of key journals. The case studies of a decision analytic model involved highlighting specific features of previously published economic studies related to generalisability and location-related variability. The case-study involving the secondary analysis of cost-effectiveness analyses was based on the secondary analysis of three economic studies using data from randomised trials.
Results: The factor most frequently cited as generating variability in economic results between locations was the unit costs associated with particular resources. In the context of studies based on the analysis of patient-level data, regression analysis has been advocated as a means of looking at variability in economic results across locations. These methods have generally accepted that some components of resource use and outcomes are exchangeable across locations. Recent studies have also explored, in cost-effectiveness analysis, the use of tests of heterogeneity similar to those used in clinical evaluation in trials. The decision analytic model has been the main means by which cost-effectiveness has been adapted from trial to non-trial locations. Most models have focused on changes to the cost side of the analysis, but it is clear that the effectiveness side may also need to be adapted between locations. There have been weaknesses in some aspects of the reporting in applied cost-effectiveness studies. These may limit decision-makers' ability to judge the relevance of a study to their specific situations. The case study demonstrated the potential value of multilevel modelling (MLM). Where clustering exists by location (e.g. centre or country), MLM can facilitate correct estimates of the uncertainty in cost-effectiveness results, and also a means of estimating location-specific cost-effectiveness. The review of applied economic studies based on decision analytic models showed that few studies were explicit about their target decision-maker(s)/jurisdictions. The studies in the review generally made more effort to ensure that their cost inputs were specific to their target jurisdiction than their effectiveness parameters. Standard sensitivity analysis was the main way of dealing with uncertainty in the models, although few studies looked explicitly at variability between locations. The modelling case study illustrated how effectiveness and cost data can be made location-specific. In particular, on the effectiveness side, the example showed the separation of location-specific baseline events and pooled estimates of relative treatment effect, where the latter are assumed exchangeable across locations.
Conclusions: A large number of factors are mentioned in the literature that might be expected to generate variation in the cost-effectiveness of healthcare interventions across locations. Several papers have demonstrated differences in the volume and cost of resource use between locations, but few studies have looked at variability in outcomes. In applied trial-based cost-effectiveness studies, few studies provide sufficient evidence for decision-makers to establish the relevance or to adjust the results of the study to their location of interest. Very few studies utilised statistical methods formally to assess the variability in results between locations. In applied

economic studies based on decision models, most studies either stated their target decision-maker/jurisdiction or provided sufficient information from which this could be inferred. There was a greater tendency to ensure that cost inputs were specific to the target jurisdiction than clinical parameters. Methods to assess generalisability and variability in economic evaluation studies have been discussed extensively in the literature relating to both trial-based and modelling studies. Regression-based methods are likely to offer a systematic approach to quantifying variability in patient-level data. In particular, MLM has the potential to facilitate estimates of cost-effectiveness, which both reflect the variation in costs and outcomes between locations and also enable the consistency of cost-effectiveness estimates between locations to be assessed directly. Decision analytic models will retain an important role in adapting the results of cost-effectiveness studies between locations.

Recommendations for further research include: the development of methods of evidence synthesis which model the exchangeability of data across locations and allow for the additional uncertainty in this process; assessment of alternative approaches to specifying multilevel models to the analysis of cost-effectiveness data alongside multilocation randomised trials; identification of a range of appropriate covariates relating to locations (e.g. hospitals) in multilevel models; and further assessment of the role of econometric methods (e.g. selection models) for cost-effectiveness analysis alongside observational datasets, and to increase the generalisability of randomised trials.

Contents

List of abbreviations

ACS	acute coronary syndrome
ANOVA	analysis of variance
BMD	bone mineral density
CABG	coronary artery bypass graft
CEA	cost-effectiveness analysis
CEAC	cost-effectiveness acceptability curve
CHART	continuous hyperfractionated accelerated radiotherapy trial
CoHD	congenital heart disease
CI	confidence interval
CPI	Conference Papers Index
CRD	Centre for Reviews and Dissemination
CUA	cost–utility analysis
CVD	cardiovascular disease
ECG	electrocardiogram
EED	Economic Evaluation Database
GDP	gross domestic product
GLS	generalised least-squares
GM-CSF	granulocyte macrophage colony-stimulating factor
GNB	Gram-negative bacteria
GNP	gross national product
GPA	glycoprotein IIb/IIIa antagonist
GWC	good wound care
HEED	Health Economic Evaluations Database
HESG	Health Economists' Study Group
HRT	hormone replacement therapy
ICC	intraclass correlation coefficient
ICD	implantable cardiac defibrillator
ICER	incremental cost-effectiveness ratio
ICST	inhaled corticosteroid
IHEA	International Health Economics Association
INMB	incremental net monetary benefit
ISPOR	International Society for Pharmacoeconomic and Outcomes Research
ISTAHC	International Society of Technology Assessment in Health Care
IV	instrumental variable
MCMC	Bayesian Markov Chain Monte Carlo
MI	myocardial infarction
MLM	multilevel modelling
NHAR	Nottingham Heart Attack Register
NHB	net health benefit
NICE	National Institute for Clinical Excellence
NMB	net monetary benefit
NSAID	non-steroidal anti-inflammatory drug
OLS	ordinary least-squares
OR	odds ratio
PCI	percutaneous coronary intervention
PPP	purchasing power parity
PRAIS-UK	Prospective Registry of Acute Ischaemic Syndromes in the UK
PTCA	percutaneous transluminal coronary angioplasty
QALY	quality-adjusted life-year

continued

List of abbreviations *continued*

QoL	quality of life
RCT	randomised controlled trial
RTI	respiratory tract infection
4S	Scandinavian Simvastatin Survival Study
WOSCOPS	West of Scotland Coronary Pravastatin Study

All abbreviations that have been used in this report are listed here unless the abbreviation is well known (e.g. NHS), or it has been used only once, or it is a non-standard abbreviation used only in figures/tables/appendices in which case the abbreviation is defined in the figure legend or at the end of the table.

Executive summary

Background

Given the increasing need for economic evidence to inform the resource allocation decisions of a range of decision-makers and in many jurisdictions, there is interest in the generalisability of economic evaluations, that is, the extent to which the results of a study based on measurement in a particular patient population and/or a specific context hold true for another population and/or in a different context. The **context** which is the primary focus of this report is the **location** in which the study was undertaken and/or the **decision-maker** for whom the study was undertaken. The focus of this report is **economic evaluation** as applied to health services.

Aims and objectives

The aim of the project was to review, and to develop further, the methods used to assess and to increase the generalisability of economic evaluation studies.

The specific objectives were to conduct:

1. A systematic review of methods literature on generalisability relating to economic evaluation to identify factors causing variability in cost-effectiveness between locations and over time, and the extent of that variability.
2. A systematic review of methods literature on economic evaluation relating to available methods to assess variability between locations and over time.
3. A systematic review of applied economic evaluation studies undertaken alongside multilocation trials to describe how studies have assessed and reported generalisability and variability in results between locations.
4. A series of case studies involving the secondary analysis of cost-effectiveness analyses undertaken alongside multilocation trials to explore the use of multilevel modelling to assess variability in cost-effectiveness between locations.
5. A structured review of economic evaluations based on decision analytic models in the field of osteoporosis to describe how studies have made their analyses relevant to particular decision-makers/jurisdictions and assessed how results might vary across locations.
6. A case study of a decision analytic model to illustrate methods to estimate cost-effectiveness for the NHS based on data partly collected in non-UK locations.

Methods

For Objectives 1 and 2 above, **methodological** studies relating to economic evaluation in healthcare were searched. This included electronic searches of a range of databases, including PREMEDLINE, MEDLINE, EMBASE and EconLit, and manual searches of key journals. Similar methods were used for Objectives 3 and 5 to identify **applied** economic studies. The case studies (Objectives 4 and 6) involved highlighting specific features of previously published economic studies related to generalisability and location-related variability. In the case of Objective 4, the case-study was based on the secondary analysis of three economic studies using data from randomised trials.

Results

Variability in cost-effectiveness by time and place

- The factor most frequently cited as generating variability in economic results between locations was the unit costs associated with particular resources.
- Some of the most frequently cited factors are as much associated with the measurement of **effectiveness** as with cost-effectiveness (e.g. the artificial characteristics of trials and patient case mix).
- No studies were identified which explicitly considered factors causing variability in the results of economic studies over time.
- Several authors have shown important variations between locations in the volume and cost of resource use and in cost-effectiveness.

Methods to assess variability in cost-effectiveness by time and place

- In the context of studies based on the analysis of patient-level data, regression analysis has been advocated as a means of looking at variability in economic results across locations. These methods have generally accepted that some components of resource use and outcomes are exchangeable across locations whereas others are not.
- Recent studies have also explored, in cost-effectiveness analysis, the use of tests of heterogeneity similar to those used in clinical evaluation in trials.
- The decision analytic model has been the main means by which cost-effectiveness has been adapted from trial to non-trial locations. Most models have focused on changes to the cost side of the analysis, but it is clear that the effectiveness side may also need to be adapted between locations.
- The review failed to identify a major literature on variability in cost-effectiveness over time, although an emerging literature using Bayesian decision theory may be of value.

Dealing with variability by location in economic studies alongside multilocation trials

- There have been weaknesses in some aspects of the reporting in applied cost-effectiveness studies. These may limit decision-makers' ability to judge the relevance of a study to their specific situations.
- There was little use of the statistical approaches identified in the methods review to assess variability by location.
- The case study demonstrated the potential value of multilevel modelling (MLM). Where clustering exists by location (e.g. centre or country), MLM can facilitate correct estimates of the uncertainty in cost-effectiveness results.
- MLM also provides a means of estimating location-specific cost-effectiveness.
- The use of location-specific covariates in MLM can explain some of the variation in cost-effectiveness.
- An important policy issue is raised by this work: the extent to which location-specific estimates of incremental net benefit are useful to decision makers.

Use of decision analytic models to provide location-specific estimates of cost-effectiveness

- The review of applied economic studies based on decision analytic models showed that few studies were explicit about their target decision-maker(s)/jurisdictions.
- The studies in the review generally made more effort to ensure that their cost inputs were specific to their target jurisdiction than their effectiveness parameters.
- Standard sensitivity analysis was the main way of dealing with uncertainty in the models, although few studies looked explicitly at variability between locations.
- The modelling case study illustrated how effectiveness and cost data can be made location-specific. In particular, on the effectiveness side, the example showed the separation of location-specific baseline events and pooled estimates of relative treatment effect, where the latter are assumed exchangeable across locations.

Key recommendations

Economic evaluation using patient-level data

- At the design stage of a study, selection of study sites should ideally focus on those that are representative of the jurisdiction(s) for which economic data are required.
- There is value in collecting data on the characteristics of trial centres which could be used as covariates in regression models.
- The patients included in studies should reflect the normal clinical caseload, but it is important to collect a number of patient-level variables that could be used as covariates.
- Resource use data (e.g. hospital days) should be reported separately from the unit costs of those resources.
- MLM should be considered as a means of assessing the degree of clustering in cost and effectiveness data within trial locations. If clustering is extensive, MLM can reflect this characteristic at the analysis stage and generate location-specific estimates of cost-effectiveness.
- There remains an important role for sensitivity analysis in exploring the implications of variation in some parameters (e.g. unit costs and preference values).
- Reporting more information on the centres/countries in a study can assist decision-makers in interpreting the relevance of results to their situation.

Economic evaluation using decision analytic modelling

- Given the focus on a **decision**, any analysis should be clear about the specification of the decision problem and the relevant decision-maker(s) and jurisdiction(s).

- The overall analytical approach, model structure and data inputs should be appropriate to the relevant decision-maker(s).
- Where several sources of data exist for a particular parameter, these should be pooled in such a way that the uncertainty relating to their precision and possible heterogeneity (including that related to location) is reflected in the model.
- It is important to distinguish parameter uncertainty from variability or heterogeneity, where the latter is concerned with how parameter estimates vary across 'contexts'.
- Probabilistic analysis, where data inputs are incorporated as random variables, is the appropriate means of handling parameter uncertainty.
- When a model is targeted at more than one decision-maker/jurisdiction, an important aspect of the analysis is to assess the **variability** in results between locations, for example, using sensitivity or scenario analysis.

Conclusions

A large number of factors are mentioned in the literature that might be expected to generate variation in the cost-effectiveness of healthcare interventions across locations. Several papers have demonstrated differences in the volume and cost of resource use between locations, but few studies have looked at variability in outcomes.

In applied trial-based cost-effectiveness studies, few studies provide sufficient evidence for decision-makers to establish the relevance or to adjust the results of the study to their location of interest. Very few studies utilised statistical methods formally to assess the variability in results between locations. In applied economic studies based on decision models, most studies either stated their target decision-maker/jurisdiction or provided sufficient information from which this could be inferred. There was a greater tendency to ensure that cost inputs were specific to the target jurisdiction than clinical parameters.

Methods to assess generalisability and variability in economic evaluation studies have been discussed extensively in the literature relating to both trial-based and modelling studies. Regression-based methods are likely to offer a systematic approach to quantifying variability in patient-level data. In particular, MLM has the potential to facilitate estimates of cost-effectiveness which both reflect the variation in costs and outcomes between locations and also enable the consistency of cost-effectiveness estimates between locations to be assessed directly. Decision analytic models will retain an important role in adapting the results of cost-effectiveness studies between locations.

Summary of recommendations for further research

Drawing on the material in this report, it is possible to summarise some important areas for further research. As far as possible, these have been placed in priority order.

- The development of methods of evidence synthesis which model the exchangeability of data across locations and allow for the additional uncertainty in this process. These methods should relate to all parameters relevant to economic evaluation.
- Assessment of alternative approaches to specifying multilevel models to the analysis of cost-effectiveness data alongside multilocation randomised trials.
- Identification of a range of appropriate covariates relating to locations (e.g. hospitals) in multilevel models.
- Further assessment of the role of econometric methods (e.g. selection models) for cost-effectiveness analysis alongside observational datasets, and to increase the generalisability of randomised trials.

Chapter 1
Introduction

Economic evaluation

The economic evaluation of healthcare technologies (including interventions and programmes) involves the comparisons of alternative options in terms of their costs and their consequences.[1] Although a number of types of evaluation exist, the majority of applied studies have been cost-effectiveness or cost–utility studies where the differential cost of the options has been related to a range of intermediate effects or measures of health gain such as quality-adjusted life-years (QALYs).[2] Full details of economic evaluation methods, as applied to healthcare, can be found elsewhere.[1,3]

In recent years, there have been some important developments in economic evaluation. The first important development in economic evaluation has been its increasing prominence in healthcare decision-making. Although there is continued uncertainty about the role of economic evaluation studies in decision-making at the level of individual hospitals and health authorities,[4] a number of healthcare systems are now using economic evaluation to make system-level decisions about which interventions to fund from collective resources.[5] In the UK, the National Institute for Clinical Excellence (NICE) was set up with the issue of cost-effectiveness central to its mission.[6] Economic evidence has been used for some years in Australia[7] and Canada[8] to establish whether new pharmaceuticals represent a cost-effective use of the resources available to the public healthcare system. More recently, a number of European countries, other than the UK, have developed an economic dimension to the regulation of healthcare technologies, including Portugal, Sweden and Finland.[5] Even in the USA, the need to ensure efficient use of collective healthcare resources has led some health maintenance organisations to use formal economic criteria in decision-making about which interventions will cover.[9]

The second development is the emergence of new economic evaluation methods in particular areas. These include alternative approaches to handling uncertainty in the context of studies based on patient-level data (e.g. randomised trials),[10] and in decision models;[11,12] and preference-based measures of health status which link data on patients' health states, as collected in trials and similar studies, with the public's health state preferences to facilitate estimates of QALYs.[13] There remain, however, a number of important sources of controversy in the field, for example, the role and methods of productivity cost estimation[14] and how to reflect equity considerations in economic evaluation.[15]

One area of methodology on which much has been written but in which few new methods have emerged relates to the generalisability of economic evaluation. The increasing need for economic evidence to inform policy decisions, but the inevitable limits on the rate at which such studies can be undertaken and published, has raised questions about the extent to which the conclusions of a given study undertaken for one specific context hold true for others. This has also stimulated interest in new methods to assess quantitatively the extent of variability in results and to make adjustments across contexts. Examples of some of these questions are highlighted in *Table 1*.

TABLE 1 *Common questions regarding the generalisability of economic evaluation studies*

What factors within an economic evaluation might limit the generalisability of studies with respect to time and place?
Which of these factors are likely to have the greatest impact on the conclusions of the studies?
What methods have been and could be used in studies with patient-level data to quantify the degree of variability in cost-effectiveness between locations?
What methods have been and could be used to extrapolate cost-effectiveness from data collected in specific locations to others where such data were not collected?
What characteristics should an economic evaluation have in terms of overall design, data collection, analysis and presentation in order to maximise its generalisability?
What should be done at the time of the study to increase the endurance of the findings?

In the evaluation literature, generalisability has been used loosely and often interchangeably with other terms such as transferability, extrapolation and external validity. For the purposes of this study, generalisability is taken to refer to the extent to which the results of a study, as they apply to a particular patient population and/or a specific context, hold true for another population and/or in a different context.[16] In the clinical evaluation literature, issues of generalisability focus mainly on the characteristics of patients in a given study and how representative they are of a broader population.[17] The **context** which is the primary focus of this report is the **location** in which the study was undertaken and/or **the decision-maker for whom the study was undertaken**. Of course, location and patient characteristics are interrelated in that the cost-effectiveness of an intervention might vary between locations because those locations treat different types of patient. However, generalisability across locations also depends on the extent of variability in other factors such as the organisation of a healthcare system and the prices of particular inputs into healthcare. Although the focus of the study is generalisability with respect to location, there is also consideration of generalisability when the 'context' refers to different time periods.

Alternative vehicles for economic evaluation

In considering issues of generalisability in economic evaluation, it is important to distinguish two types of 'vehicle' for economic evaluation – studies collecting patient level data and decision analytic models. A large number of economic evaluation studies are based on patient-level data on resource use and outcomes which are taken from a single study. Economic evaluation alongside a single randomised trial is probably the most prevalent example of this type of study. In contrast, the decision analytic model estimates the cost-effectiveness of alternative interventions by synthesising aggregated data from a range of different sources including single trials, meta-analyses of trials, observational studies and surveys.[18] Decision models represent an important approach to economic evaluation as they provide a framework within which all forms of uncertainty can be explicitly quantified and its implications assessed.

These different approaches to economic evaluation raise different questions concerning generalisability. The trial-based economic evaluation seeks to estimate costs and effects in a sample of patients and locations which it hopes will be representative of a broader context. Hence key questions relating to generalisability in trial-based economic evaluation include:

- How representative is the trial sample of the patient population from which it is drawn (i.e. in the recruiting centres)?
- How representative of the relevant patients outside the recruiting centres is the trial sample?
- How representative of all centres in a particular jurisdiction are those recruiting into a trial?
- In multicentre/multinational trials, how much variability in costs, effects and cost-effectiveness was there between locations?

The use of decision modelling as a vehicle for economic evaluation typically sets out to provide information to particular decision-makers or jurisdictions. However, to the extent that these jurisdictions include a range of contexts (e.g. patient subgroups and locations), many of the questions raised above, for trial-based studies, also apply to models. There are, however, additional questions relating to generalisability and variability in models, and these include:

- Have the analysts been explicit in defining the decision maker(s) or jurisdiction(s) which their model is seeking to inform?
- Are the input parameters in that model the most appropriate for those particular decision-makers/jurisdictions?
- To what extent have the analysts sought to explore how robust their conclusions are to alternative input parameters that might apply to other decision-makers/jurisdictions?
- Have the analysts sought to adjust those input parameters, which have been estimated in different locations, to be appropriate for the target jurisdiction? What methods have been employed for such adjustment?

The importance of generalisability

A greater understanding of generalisability in economic evaluation has potential benefits for three main groups: investigators, decision-makers regarding service provision and research funding bodies. From the investigators' viewpoint, there is value in information regarding the elements of an economic evaluation which are likely to vary across locations in such a way as to alter the conclusions of a particular analysis. Investigators

will also benefit from the identification and development of methods to assess the extent of variability in cost-effectiveness between locations and its implications for results and methods to extrapolate the results of a given study to locations for which the analysis was not originally intended.

Decision-makers concerned with service provision face two important questions in assessing available economic studies. First, are the methods employed in the study appropriate and are the results valid? Second, if the results are valid, would they apply to the populations and settings for which the decision maker has specific policy responsibilities. Given the plethora of different decision-makers at different levels of various healthcare systems, it will never be possible for each to mount their own economic evaluation from scratch, and most will need to interpret available studies. Therefore, the sorts of questions outlined above for both trials and models will need to be addressed by this group.

In general terms, whether funding an economic evaluation is worthwhile for the research funder will depend on the degree of uncertainty in existing cost-effectiveness evidence and the implications of that uncertainty.[12] Part of the process of quantifying the uncertainty in existing evidence is the extent to which available data have been generated in the relevant jurisdiction and, if not, extrapolating from those data. This process is often undertaken informally based on judgement in the face of evidence reviews, but analytical methods are able to inform this process with the additional uncertainty associated with taking evidence from other locations being explicitly reflected in those methods. In reviewing trial proposals, funders will also be concerned with the likely generalisability of the study for their relevant jurisdiction; for decision models, the appropriateness of the parameter inputs for their jurisdiction will be a concern. Hence the sort of questions outlined above will need to be addressed, but at the stage of design rather than after the analysis.

The study's focus

The focus of this report is **economic evaluation** as applied to health services. However, given that the cost-effectiveness of an intervention is closely related to its effectiveness, there is inevitable overlap between issues of generalisability in economic analysis and those relating to clinical evaluation. This project did not, however, review the wider literature on generalisability in clinical trials or epidemiology. In part, this is because the NHS Health Technology Assessment Programme has already funded work in these areas.[19] A number of important methodological topics, which have important implications for evaluative health services research in general are not, therefore, formally reviewed in detail here. However, we did seek to acquire a broad understanding of some of these methods by talking to experienced researchers in this field. These methods include, for example, comprehensive cohort designs,[20] multilevel models[21] and selection models.[22,23] However, some of these concepts and research areas have now been applied to economic evaluation and, by virtue of this, they are included in specific parts of this report.

The project also did not look at literature outside health and healthcare. In principle, issues of generalisability and transferability are relevant in areas such as operations research and transport and environmental evaluation, but the extent of literature searching necessary to understand and present this literature systematically was considered too great to be part of this project. Again, we discussed the issues with experienced researchers in these fields to get a broad understanding of the insight these disciplines might provide. Although we cannot discount the possibility that important and, as yet, unidentified concepts have been developed in these areas which would have potential value in economic evaluation in health, our discussions with experts suggested this was unlikely to be the case.

Aims and objectives

The aim of the project reported here was to review and to develop the methods used to assess and to increase the generalisability of economic evaluation studies.

The specific objectives of the report are detailed below. Each objective has been addressed in one of the following chapters, which are also indicated.

- A systematic review of economic evaluation literature on generalisability to identify factors causing variability in cost-effectiveness between locations and the extent of that variability (Chapters 2 and 3).
- A systematic review of economic evaluation literature on available methods to assess

variability, and to extrapolate, between locations (Chapters 2 and 4).
- A systematic review of applied economic evaluation studies undertaken alongside multicentre randomised trials to describe how studies have assessed and reported generalisability and variability in results across locations (Chapter 5).
- A series of case studies involving the secondary analysis of cost-effectiveness analyses undertaken alongside multilocation randomised trials to explore the use of multilevel modelling to assess variability in cost-effectiveness between locations (Chapter 6). These case studies were based on multilocation trials for which we had patient-level data available.
- A systematic review of economic evaluations based on decision analytic models to describe how studies have made their analysis relevant to a particular decision-maker/jurisdiction and assessed how results might vary across location (Chapter 7). This review is based on a particular clinical area – osteoporosis – which we considered exhibited a number of valuable features for the exercise, for example, the large number of cost-effectiveness models in the area and the likely variability in some parameters between countries and other locations.
- A case study of a cost-effectiveness study based on a decision analytic model to illustrate available methods to estimate cost-effectiveness for the NHS based on extrapolation from data collected in non-UK locations (Chapter 8). This case study was selected as it raises a number of issues about the transferability of data between locations.
- A set of recommendations regarding the design, execution and reporting of economic evaluation studies to increase their generalisability and to facilitate an assessment of variability in cost-effectiveness across locations (Chapter 9).

Chapter 2

A systematic review of methodological literature: review methods

Chapters 2–4 detail a systematic review of health economics literature relating to generalisability in economic evaluation in the healthcare field.

The objectives of the review were:

- To identify factors that affect the generalisability and transferability of economic evaluation results.
- To identify published economic evaluations which have sought to estimate variability between locations and over time.
- To identify and review proposed methods to increase the generalisability of economic evaluations.
- To identify previously published studies that illustrate the application of proposed methods to increase the generalisability of economic evaluations.
- To identify on the basis of the review methods that require dissemination and areas where further methods need to be developed.

The 'dimensions' of generalisability that were of interest were principally those relating to geographical setting – that is, the extent of variation in economic results between geographical locations. However, issues relating to variability in results across time were also considered.

This chapter outlines the review methodology, Chapter 3 discusses conceptual papers on factors affecting generalisability and empirical papers estimating variability and Chapter 4 discusses papers dealing with methods to analyse variability or increase generalisability.

Overview

A systematic and replicable search strategy was devised to search for relevant articles under the four categories below:

- conceptual papers on factors affecting the generalisability of results in economic evaluation studies
- conceptual papers on methods to increase generalisability in such studies
- empirical papers estimating variability in economic evaluation results between locations or across time
- empirical applications of methods to increase generalisability in economic evaluation.

It was expected that the majority of the key papers would be identified in methods journals specific to the field of economic evaluation and health technology assessment. The search was not, however, limited to these sources, and included publications in the epidemiological, statistical and policy fields, although here papers were sought which related specifically to economic evaluation. Journal articles comprised the main body of the literature but all sources, including books, reports and conference papers, were considered for inclusion.

Inclusion criteria

The inclusion criteria that were relevant to the review are defined formally below:

1. Conceptual papers on factors affecting generalisability
   (a) Including 'think pieces' and reviews of methods papers.
   (b) Papers considering factors relating to one or both of the two dimensions (time or location) of generalisability.
   (c) Factors which affect cost-effectiveness through an impact on resource use and/or outcomes of healthcare interventions.
2. Conceptual papers on methods to increase generalisability
   (a) Papers considering methods to increase one or both of the two dimensions of generalisability (time or location).
   (b) Methods seeking to increase generalisability in cost-effectiveness estimates through an effect on resource use and/or outcomes of health care interventions.
3. Empirical papers estimating variability in results between locations or across time

(a) Papers quantifying variability by location or time.
(b) Empirical papers where it was judged that variability by time or location was estimated as one of the primary objectives or primary or principal analyses, or for purposes other than assessment of the uncertainty of the data (i.e. sensitivity analysis).
(c) Excluding applied economic evaluations that looked at variability as a sensitivity analysis other than by location or over time.
(d) Excluding studies which looked at variability by patient subgroups without explicitly linking this to time or place.
(e) Excluding qualitative descriptions of variability.

4. Empirical application of methods to increase generalisability in economic evaluation
   (a) Empirical papers with the objective of illustrating, testing or appraising a method to increase generalisability in cost-effectiveness by location or time.
   (b) Papers based on actual data rather than hypothetical estimates.
   (c) Excluding applied economic evaluations that assessed generalisability as a secondary consideration (e.g. sensitivity analyses)

Two further exclusion criteria were also employed:

- Papers prior to 1985. Scoping searches indicated that it was very unlikely that relevant economic evaluation literature existed prior to this date.
- Non-English language papers.

Search strategy

Electronic databases

The search strategy used a variety of approaches. The main part was undertaken using electronic databases, but references were also obtained by other methods, such as handsearching journals, recommendations from the expert panel and citation searches. Once identified, references were managed, and duplicate entries identified, using an EndNote Version 5.0 database.

A set of search terms was identified and used to search a selection of electronic bibliographic databases. The searches aimed to find papers relating to generalisability over location or time in relation to economic evaluations. The searches were restricted to 1985 onwards, and no language restrictions were applied at the search stage to enable an approximate estimate of the size of the non-English language literature to be obtained.

The search strategies were updated periodically and, in total, three sets of searches were conducted. All the references were imported into an EndNote library and were de-duplicated. The databases searched included a range of health-related databases, PREMEDLINE, MEDLINE and EMBASE, and also some specialist economics databases: EconLit, Economic Working Papers Database, Health Economic Evaluation Database (HEED) and the NHS Economic Evaluation Database (NHS EED). The NHS EED database was searched using the CRD administration database, which is more current and broader than the public version of NHS EED. In addition, the in-house catalogue of the Centre for Reviews and Dissemination/Centre for Health Economics Information Service (CRD/CHE) was also searched. The Conference Papers Index (CPI) was searched for conference presentations to widen the search for unpublished material. The search strategies for all of the databases are contained in Appendix 1.

Handsearches of journals

Selected journals were also searched by hand from 1985 onwards. The following journals were included: *(British) Journal of Medical Economics*, *Controlled Clinical Trials*, *Health Economics*, *Health Economics in Prevention and Care*, *Health Policy*, *International Journal of Technology Assessment in Health Care*, *Journal of Health Economics*, *Medical Decision Making and Pharmacoeconomics*, *Value in Health*.

Citation searches

Citation searches were carried out on 13 key articles using the Social Science Citation Index (SSCI) via Web of Science. The articles selected were those thought to be the most commonly cited ones discussing factors affecting generalisability and relevant methodological issues.[16,24–34]

Bibliographies of included articles

The bibliographies of all papers fulfilling the inclusion criteria were examined and potentially relevant papers were acquired.

Articles from experts

At meetings of the project's Expert Advisory Group, members were asked to identify further potentially relevant articles.

Articles from conferences

In addition to the CPI, some key conference proceedings and abstract books were searched

from 1996 to 2001. The following conferences were included: International Society of Technology Assessment in Health Care (ISTAHC), International Health Economics Association (IHEA), International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and the UK Health Economists' Study Group (HESG).

Sifting and data extraction

Titles and, where available, abstracts, from all the searches were sifted independently by two researchers to identify potentially relevant papers; any difference of opinion was settled by discussion. Hard copies of all potentially relevant papers were obtained. These were then studied against the inclusion criteria set out above, and a final set of papers for data extraction was identified.

Once a paper was identified as being relevant, it was classified into one or more of the four categories above. Information from included papers was extracted by one Research Fellow (FSP) using a review pro forma which had been developed based on the objectives of the project (see Chapter 1) and the researchers' prior knowledge of the area, and piloted using a sample of eight papers. The items in the pro forma were the column headings in the tables in Appendix 2.

TABLE 2 *Breakdown of sources of articles identified and included from electronic databases*

Database	Host	Dates covered[a]	Identified[b]	Included
MEDLINE	Silverplatter/ARC	1980–Nov. 2001	1108	21
PREMEDLINE	PubMed	Up to 4 Feb. 2002	121	0
EMBASE	Silverplatter/ARC	1980–Jan. 2002	432	17
EconLit	Silverplatter/ARC	1969–Nov. 2000	1542	1
HEED	CD-ROM	Jan. 2002 issue	248	16
Economic Working Papers Database	http://econwpa.wustl.edu:80/months/hew http://ideas.uqam.ca/ http://netec.mcc.ac.uk/BibEc.html	Up to 6 May 2002	5	0
CRD/CHE Catalogue	CAIRS	Up to 5 Feb. 2002	26	1
CPI	Dialog	1973–Feb. 2002	274	0
NHS EED	CAIRS	1994–5 Feb. 2002	281	9

[a] Papers with publication dates from 1985 onwards were sought, but the databases were searched from as far back as possible.
[b] Prior to de-duplication. After de-duplication, there were 3353 records.

TABLE 3 *Breakdown of sources of articles identified and included from handsearches and conferences proceedings/databases*

Journal/conference	Dates covered	Included
Journal		
British Journal of Medical Economics	1993–2001	0
Controlled Clinical Trials	1990–2001	3
Health Economics	1992–2001	3
Health Economics in Prevention and Care	2000–2001	1
Health Policy	1993–2001	2
International Journal of Technology Assessment in Health Care	1990–2001	6
Journal of Health Economics	1990–2001	0
Medical Decision Making	1997–2001	1
Pharmacoeconomics	1992–2001	8
Value in Health	1998–2001	1
Other	1990–2001	7
Conferences		
UK HESG	1998–2001	2
ISTAHC	1997–2001	0
ISPOR	1998–2001	4
IHEA	1999–2001	6

Papers identified

The electronic search generated 4037 references in total which, after duplication, amounted to 3353 unique papers. Of these, 304 articles were retrieved, of which 65 were considered to be consistent with the inclusion criteria and were formally included in this review. In addition, 31 articles were identified through handsearches, 12 papers through a review of conference proceedings/databases and one article from the biographies of included papers. Citation searches generated 318 articles from the 13 key references. When these citations were reviewed, all the relevant references were found to have been previously identified. In total, 109 articles were included in the review. Details of the sources of identified and included articles are provided in *Table 2* (for electronic databases) and *Table 3* (for other sources).

Chapter 3

A systematic review of methodological literature: factors affecting variability and empirical papers estimating variability in economic evaluations

This chapter discusses papers that explore variability in one of two ways: first, those that consider variability in **conceptual** terms by identifying factors that might affect cost-effectiveness results between geographical locations or over time; and second, several papers report empirical studies that attempt to estimate the variability in cost-effectiveness results, either between geographical locations or over time. In both cases most of the literature relates to variability by geographical location.

Conceptual papers on factors affecting generalisability

Basic results

The papers considered in this section are those that considered factors that are likely to generate variability in cost-effectiveness results between locations or across time. A number of authors have suggested a whole range of potential factors that might influence geographical variability in cost-effectiveness results. Of the 109 papers included in the overall review, 36 studies consider sources of variability in conceptual terms. Appendix 2A summarises each of these papers in terms of their objectives, whether or not they used formal systematic review methods, whether they were focused on a particular vehicle for economic evaluation (e.g. trial or decision model), whether the paper concentrated specifically on issues relating to the generalisability of economic studies in the NHS, which factors the papers identified as potentially generating variability and which dimension of generalisability was considered (i.e. by location or time). The table shows that none of the studies used formal systematic review methods as part of their reviews; 50% of studies focused on issues of variability relating to trial-based economic evaluation, 3% on issues concerning economic studies based on models, with the remainder (47%) not focusing on any particular framework. All of the papers considered factors affecting geographical variability in economic studies and, of these, 32% also provided some discussion of factors impacting on variability over time in the results of economic studies.

Potential causes of variation in cost-effectiveness between locations

Papers looking at potential sources of variation by location are summarised in Appendix 2A. The factors identified in the papers are summarised in *Table 4*. It can be seen there is a fair degree of consistency between authors. The most frequently mentioned factor is the absolute or relative prices of resources, that is, the unit costs used to translate measures of physical resource use into costs in monetary terms. This is clearly an economic variable in the sense that it is used in economic evaluations, but not in clinical assessments (e.g. effectiveness analyses from randomised trials). However, it is interesting that the next most frequently cited factors are likely to be relevant to clinical evaluation in addition to economic evaluation: clinical practice variations; the artificial conditions existing in the centres from which study patients are drawn, when compared to routine practice; and the skills and experience of staff. Other factors frequently referred to were variations in resource use between centres, demography and differences in the organisation of the healthcare system.

The key factors considered important as a source of geographical variability are now discussed in more detail. These have been arranged into four groups: patient factors, clinician factors, healthcare system factors and wider socio-economic factors.

Patient factors (B, D, H, L in Appendix 2A)

In economic evaluation generally, arguably the most important source of variation in cost-effectiveness is between subgroups of patients defined in terms of demographic and clinical factors. For example, age can have an impact on cost-effectiveness.[35] For this reason, patient subgroup analysis presents an important challenge to economic evaluation methodologically and in terms of policy.[6] This

TABLE 4 *Conceptual factors that can cause variability in economic studies: factors cited by authors*

Factor	Definition	References
Absolute/relative costs	Unit costs/prices of inputs into healthcare	1, 16, 27, 30–32, 39, 41, 45, 47, 50, 51, 53–60, 109, 110
Artificial study conditions	Research environment versus routine practice	16, 35, 37, 38, 40, 46, 50, 60–63, 111, 112
Capacity utilisation	Level of utilisation of inputs into healthcare	1, 16, 38, 40, 51, 53
Case mix	Clinical and socio-demographic characteristics of patients undergoing treatment	16, 38, 46, 50, 51, 58, 63, 113
Clinical practice variation	Variation in how healthcare is delivered	1, 16, 27, 31, 32, 36, 39–41, 54, 56–59, 61–64, 110
Compliance	Adherence to treatment regimen	39, 40, 44–47, 50, 58, 114
Culture/attitudes	As affecting clinical practice	39, 45, 62, 63
Demography	Patient non-clinical characteristics, e.g. sex, age	1, 31, 35–37, 39–41, 43
Disease interaction	Association of primary disease with risk factors, other morbidity/mortality	38, 46, 50
Economies of scale	Greater levels of 'production' leads to lower costs	38, 40, 51, 59
Epidemiology	Incidence/prevalence of disease	1, 30, 31, 39, 41, 51
Exchange rates	Conversion rate of different currencies	50, 54
Geographical setting	Location such as country, type of facility	37, 39, 50, 53, 59, 62, 109, 111, 112
Health state valuations	Individuals' preferences for particular levels of health	16, 50
Healthcare resources	Inputs into health delivery, e.g. personnel, equipment	1, 27, 31, 32, 38, 40, 41, 45, 47, 53, 54, 58, 63, 64, 113
Healthcare system	Regulatory and organisational infrastructure	1, 32, 38–40, 45–47, 50, 51, 55–57, 59, 110, 113
Historical differences	History of organisation/practice	46
Incentives	Financial and other factors which affect individuals' and organisational behaviour	1, 27, 31, 41, 50, 51
Industry-related bias	Sponsor influence on study results	37, 56
Joint production	Inputs into healthcare delivery are shared between different units/departments	38, 53
Opportunity cost	Health benefits forgone by use of a resource in a particular way	50
Perspective	Viewpoint of economic analysis	27, 36, 51
Skills/experience	Level of training and experience of health professional	36–40, 45, 50, 53, 55, 58–60, 64, 112, 113
Technological innovation	Advancement of technology/practice	36, 47, 53
Timing of economic evaluation	Stage of conduct of study in the development of the technology	43, 57
Treatment comparators	Available treatment options	1, 27, 30, 37, 43, 50, 57

important source of patient-level variation feeds through to centre or country variations in cost-effectiveness if these subgroups of patients are not evenly distributed between locations. This can partly be explained in terms of demography, where local populations differ with respect to age structure, life expectancy, gender, genetic make-up and lifestyle. These characteristics will affect the presentation of disease and thus the cost-effectiveness of healthcare programmes.[31,36–42]

Related to demographics, variation between locations may also be due to differences in epidemiology, that is, differences in the incidence and prevalence of diseases and subgroups of diseases. This can create different case-mixes between locations which may affect the cost-effectiveness of treatment.[30,31,39,41] Although it is likely that there are differences between countries, differences may also exist within countries. Drummond and colleagues[1] suggest that preventive programmes are more likely to be cost-

effective in populations where the incidence of the 'target' disease is high.

Even if demographics and epidemiology are similar between locations, the characteristics of those patients who are treated may vary. This difference in **case mix** can be due to different thresholds being used to judge which patients undergo treatment. Within regions of individual countries, differences between centres in terms of case mix can also be due to specialisation – for example, some centres may choose to focus on more complex cases.

One source of variability relates to the differences between those centres which typically undertake research studies, such as clinical trials, and those which do not. This is a special case of variability by location in that it implies that measurements undertaken in those centres (e.g. clinical effects and resource use) may not be a realistic prediction of what might emerge should the relevant interventions be undertaken in other (i.e. non-study) centres. These differences exist at the level of the patient, the clinician and the healthcare system. At the patient level, many studies (particularly pivotal clinical trials which are undertaken to secure regulatory licences for new pharmaceuticals) recruit particular types of patients to further the objectives of the study, rather than recruiting a group which is broadly representative of the relevant patient population.[43] For example, regulatory trials often include more high-risk patients in order that the baseline risk of a given clinical event is higher, which gives more scope for a greater absolute treatment effect.[19,43] Furthermore, these trials also often exclude patients with co-morbidity.

One characteristic observed in clinical research which may not be replicated in routine practice is patients' levels of compliance. In general, compliance is expected to be higher in the study environment where patients are closely supervised. Both the treatment effect and cost can vary with compliance in the case of a drug therapy, and it is arguably the most important element responsible for differences that exist between the effectiveness and efficacy of an intervention.[44] The consequence of non-compliance is likely to be increased resource utilisation (such as hospitalisation costs, prescription changes, diagnostic tests, consultations and more frequent visits) owing to reduction in efficacy and associated risk in therapeutic failure. If compliance is higher in research than routine practice, then economic studies based on clinical trials may underestimate costs.[38–40,45–47] Differences in education levels and cultural factors may also directly affect compliance and hence could impact upon the implementation of different strategies for treatment and their cost-effectiveness. For example, in certain countries people have objected to residual house spraying with an insecticide for malaria eradication because of factors such as the inconvenience. As a result, the WHO now recommends spraying in certain circumstances, therefore causing the cost-effectiveness to differ from settings of more general use.[39]

Clinician factors (B, E, W in Appendix 2A)

Clinicians can influence the effectiveness, cost and cost-effectiveness of interventions. This 'clinician effect' is probably less likely to manifest itself in terms of the cost-effectiveness of pharmaceutical therapies, where the intervention is usually standardised and the clinician has less opportunity to influence the process of care. In other types of interventions – for example, surgical treatments – the effect of clinician is an integral part of the treatment. With such interventions, variation in the skill and experience of the clinician can have a marked effect on the process of care and the outcomes. This has also been referred to as 'practice style' – the concept that it is the physician's habits and beliefs about the efficacy of particular forms of care.[48] However, the clinician effect is typically not easy to quantify within clinical or economic evaluations. Hence observational data have been the focus of this type of research and the results have been contentious.[49] To the extent that centres and/or countries differ in respect of factors such as the skill and experience of their clinicians, such a 'clinician effect' may manifest itself in terms of variations in cost-effectiveness between locations.

In part, variation in how clinical staff perform can be due to the fact that healthcare systems differ in terms of the incentives that they offer to staff.[1,27,31,41,50,51] For example, in a fee-for-service environment, physicians are paid a fee per item of service, and hospitals are reimbursed by the number of cases in each category treated. It has often been suggested that physicians operating under a fee-for-service system are more likely to generate extra demand for their services, whereas those paid by salary or capitation are more likely to limit demand.[52]

Health care system factors (A, B, C, E, P in Appendix 2A)

Regardless of the characteristics of patients or clinicians, there remain numerous differences

between countries and centres in terms of healthcare delivery. As shown in *Table 4*, variation in prices is the most frequently cited source of geographical variability in economic studies. Economic theory suggests that there are a number of elements that influence the cost of a good or service. These include the technology involved in production, the rate of substitution between labour and capital, the types and cost of resource inputs (e.g. labour and capital) used in production and the overall level of productive output.[53] Because these and other elements are likely to vary between institutions and geographic locations, it is natural that variations exist in unit costs. Although it might be expected that this variation would be more pronounced internationally, the economic factors that give rise to variability in prices may also exist between centres (e.g. hospitals) within countries.[16,27,30–32,39,41,45,47,54–60] This within-country variation is likely to be particularly pronounced in large and economically heterogeneous countries.

Variation in unit costs, without accompanying differences in the actual resources used, would be expected to affect only those study parameters relating to costs. It is unlikely that parameters related to health-related outcomes would be affected by differences in unit costs, unless they lead to a different mix of inputs. However, it is possible that differences in general and relative price levels for healthcare resources may affect the ranking of alternatives in a given economic evaluation with likely consequences for policy decisions.

Clinical practice and conventions are known to differ widely both within and between countries.[27,31,32,36,39,40,54,56,57,59,61–63] One aspect of this variation relates to the healthcare system and the local organisational and/or service delivery characteristics. For example, specialised medicine is predominantly hospital-based in the UK, whereas the French healthcare system has many community-based specialists. Variations in practice patterns can affect the presentation (through differential work-up of patients and differential treatment thresholds) and outcomes of the disease, and in turn can affect, or be affected by, the relative costs of care in different countries. Another aspect of this source of variability relates to the distribution and availability of healthcare facilities, which in turn leads to differences in clinical practice since physicians may face different ranges of treatment options. Such differences will lead to variations in the choice of appropriate comparators in economic studies, and could lead to results that are very different, either in terms of effectiveness or costs.

Some important research on the relationship between hospital characteristics and costs has been undertaken. In the context of economic evaluations comparing peripheral blood stem cell transplantation with autologous bone marrow transplantation as support for high-dose chemotherapy, Neymark and Rosti examined whether systematic differences in practice could be explained by differences in factors such the type of institution, the centre's cumulative experience in the field of interest or the budget mechanisms of hospitals.[64] Based on a questionnaire sent to 162 centres (60% hospitals, 14% cancer centres, 26% general hospitals), considerable variations were observed with respect to all aspects of patient management and technical procedures investigated. It was concluded that economic evaluations in the area cannot be generalised from one setting to another without careful examination of the procedures and strategies followed in each setting.

An important reason for variation in cost (and perhaps cost-effectiveness) between centres and countries is variation in resource use. For example, both Postma and colleagues[65] and Rhodes and colleagues[66] found variations in average length of stay when an identical treatment was implemented in similar populations in several countries simultaneously. In part, this variation is a result of other factors discussed in this section such as clinical practice variations. It can also be a result of differences between centres/countries in the price (unit cost) of resources. For example, if the price of expert medical staff is extremely high relative to other staff, decisions might be made to use less expensive nursing staff to undertake all or some of the duties which medical professionals undertake in other centres or healthcare systems. This potentially complex relationship between the price of particular resources and the volume with which they are used in producing healthcare is central to understanding the constraints on generalisability in healthcare and attempting to assess variability between locations.

Other factors that might result in variability in the type and volume of resources used between locations are heterogeneity in training and education and capital expenditure in previous years. There is evidence to suggest that physicians with more years of experience turn out to be significantly more costly, with differences in costs much larger than could be expected based on salary differentials alone.[38]

An important influence on the pattern, volume and cost of resources is the capacity within a healthcare provider (system or centre) and the scale at which it operates.[16,38,40] For example, differences in the intensity with which a given fixed resource (e.g. a hospital or an item of equipment) are utilised will result in the variation in the cost per patient treated. A report by the Nuffield Institute for Health concluded that economies of scale are more likely to exist in acute hospitals with 100–200 beds, whereas diseconomies of scale are likely to exist in hospitals with more than 300–600 beds.[67] Consistent with this, Cowing and colleagues have shown that in small hospitals (fewer than 100 beds), there is little evidence of significant economies of scale.[68]

The total volume and value of resources available to deliver healthcare will vary between and within countries. For example, in 1999 the UK devoted 7% of its gross domestic product to healthcare compared with 13.7% in the USA, 10.5% in Germany, 9.5% in Canada and 8.4% in Italy.[69] This source of variation will impact on a number of the factors already considered, such as variation in the type of resource used for particular interventions, the prices of resources and potentially the skill and experience of clinicians. Another implication is that there will be variation by location in the willingness to provide additional resources for a given unit of health gain. For example, if a new intervention for lung cancer becomes available which has a marked effect on health outcomes but at significant additional cost compared with standard treatment, less well resourced systems/centres will have to give up more of what they currently provide to patients across all diseases to afford the new cancer intervention compared with systems/centres with greater resources. In other words, the **opportunity** cost of taking on new healthcare interventions will be higher in less well resourced locations, to the extent to which they may be unwilling to make the investment. As O'Brien has noted, there is no obvious reason why all countries (or indeed locations within countries) should value health relative to other goods at the same rate.[50] This will show itself in variations in funding decisions between locations on the basis of the same cost-effectiveness data from studies and this, in turn, will feed into variation between locations in terms of the appropriate comparators for future studies.

Finally, the variation in the results of economic studies undertaken in research settings and those that would emerge in routine practice partly relates to the level of the healthcare system. In other words, some healthcare systems are more likely to engage in the sort of research (e.g. clinical trials) which will feed into economic studies than others. However, the evidence with respect to how research settings influence unit costs is ambiguous. Soderlund and colleagues discovered that teaching status and higher labour input had little impact on costs,[70] but regression studies have found higher costs in teaching hospitals.[71]

Wider socio-economic factors (G, O in Appendix 2A)

A number of the factors referred to in *Table 4* relate to variation between locations in terms of more general socio-economic factors than the characteristics of their patients, clinicians or heathcare system. One factor relates to the willingness of a region/country to devote resources to healthcare, and this feeds into a system's ability to fund new interventions as discussed in the last section. Another broader factor which may lead to variation between locations in the results of economic studies is the health-related preferences of the population such as those reflected in health state utilities used to calculate QALYs.[50] To the extent that the cost-effectiveness of an intervention is assessed using the preferences of a sample of the wider community rather than patients, any variation in the preferences between countries/regions may be reflected in the results of studies. There has been relatively little exploration of the extent to which health state preference values vary by geographical location and no formal analyses were identified in the systematic search. However, the research to date on health state 'utilities' suggests that the mean values for different health states do not vary greatly between locations.[72] Although the incorporation of a population's equity-related attitudes into applied economic evaluations currently remains an area of methodological, rather than applied, research, it is possible that populations will differ in the attitude towards whether more weight should be given to particular subgroups of the population.[73]

Factors affecting variation over time

The factors that may lead to variability in the results of economic studies across time have been explicitly considered far less frequently in the methods literature. Appendix 2A indicates that only 11 (32%) of the papers included in the review referred to factors which may explain this type of variation. However, it should be noted that most of the factors discussed in the last section relating to geographical variation in the results of economic studies can also vary over time. For example, the results of an economic evaluation of

a given intervention undertaken 10 years ago may vary from a similar study undertaken today because of changes in factors such as prices, resource use, patient case mix and clinical practice. In other words, factors that can be used to explain variation across locations and time are very much interrelated. Some additional factors identified in the literature are considered below.

Learning effects (W, X in Appendix 2A)

In microeconomics, the average cost per unit produced tends to decrease with increased production volume due, in part, to the learning curve whereby efficiency of production increases with experience. In the field of healthcare, learning effects can relate to increases in skill levels which take place over time and may be associated with individual clinical staff, collective experience of the organisation or technical experience with the design and application of more complex technologies such as medical devices.

Costs and outcomes of recently introduced health technologies, compared with existing health technologies, may be affected by these learning effects.[36,38,40,45,47,50,53,55,58,62] Over time, the new technology will be increasingly mastered and this can be reflected in terms of costs and outcomes. However, the direction of change in costs and outcomes as a result of the experience may be difficult to predict. For a given patient case mix, costs may be reduced and outcomes improved as a result of increased experience and technical knowledge. For example, as surgeons become more experienced with new procedures, they are likely to undertake them more quickly and with fewer adverse events. However, there may also be a tendency to use the technology on a different mix of patients, which may include more complex cases (see below). A review of learning effects in healthcare has recently been published, which discusses some of the potential effects on costs and outcomes of healthcare technologies.[74]

Treatment population (D in Appendix 2A)

Changes over time in the characteristics of patients undergoing treatment can lead to changes in the estimation of cost-effectiveness of particular technologies. First, the introduction of a healthcare technology may attract patients with a different severity of illness than the patients involved in the original research study. Second, the indications for patients eligible for treatment may expand over time, and it may be that the first patients to be treated have a higher capacity to benefit than the patients treated later. This changing threshold to intervention has been observed in the field of minimal access surgery – for example, endometrial resection in menorrhagia and laparoscopic cholestectomy.[75,76] Third, the protective immunity of populations may decline over time, as in the case of malaria where mosquitoes develop resistance to insecticides over time, causing the cost-effectiveness of treatments to change.[39]

Timing of evaluation in lifecycle (Y in Appendix 2A)

One of the factors affecting cost-effectiveness over time is the timing of evaluation. *Ceteris paribus*, the earlier an evaluation can be conducted within a technology's life cycle, the greater is the likelihood the evaluation may affect practice.[57] Once a decision-maker (e.g. physician or health policy maker) has decided to adopt a new technology, the potential for an evaluation to change such a decision may be more limited.[46]

Comparators (Z in Appendix 2A)

Scientific innovation is likely to generate new healthcare technologies. One effect of this is that the appropriate comparators, against which an intervention of interest might be assessed, will change. This can have the effect that the cost-effectiveness of the technology against one or more of the mutually exclusive alternative options which could be adopted for a particular patient group can alter over time. For example, in the field of rheumatoid arthritis over the last decade, treatments have progressed from various combinations of non-steroidal anti-inflammatory drugs (NSAIDs) and analgesics (pain-relieving drugs) to corticosteroids, disease-modifying anti-rheumatic drugs (DMARDs) and biologics such as anti-TNFα. Hence what was considered an appropriate 'standard care' comparator 10 years ago may no longer be so.

Empirical studies estimating variability in economic studies

Studies looking at geographical variability

A number of empirical studies provide evidence that the factors considered above and summarised in *Table 4* can affect the cost-effectiveness of interventions. Of the 109 references included on the database, 33 studies are relevant for inclusion in this category. Full details of these studies can be found in Appendix 2B. Also, *Table 5* summarises

TABLE 5 *Empirical studies looking at variability: factors cited by authors causing variation (the same list of factors is provided as in* Table 4*, although not all have been assessed empirically)*

Factor	References
Absolute/relative costs (prices)	31, 34, 56, 78, 80–84, 87–106, 115, 116
Artificial study conditions	34, 56, 80, 88, 99, 107
Capacity utilisation	
Case mix	89
Clinical practice variation	31, 56, 78–81, 83, 84, 87, 89–103, 105, 106, 115, 116
Compliance	31, 98
Culture/attitudes	98, 102
Demography	79, 99–101, 104
Disease interaction	
Opportunity cost	31
Economies of scale	
Epidemiology	100, 101, 106, 107
Exchange rates	
Geographical setting	31, 34, 56, 78–83, 87, 89–100, 102, 103, 105, 106, 115, 116
Health state valuations	
Healthcare resources	78, 79, 84, 96, 98, 99, 102, 115
Healthcare system	56, 89, 96, 107
Historical differences	
Incentives	
Industry-related bias	56
Joint production	
Perspective	56
Skills/experience	96, 98, 107
Technological innovation	107
Timing of assessment	107
Treatment comparators	103, 107

the sources of variation that have been explored empirically. The table shows that empirical studies indicate that geographical variation in unit costs (prices) is most frequently identified as impacting on the results of economic studies; variations in clinical practice, geographical setting and availability of healthcare resources have also been shown to be important.

Table 6 provides a summary of some of the other characteristics of the empirical studies. In terms of clinical areas as categorised by the National Library of Medicine, 'bacterial infections and mycoses' and 'cardiovascular diseases' have been the most frequently studied. Cardiovascular diseases have been of particular interest to analysts because of the considerable economic implications imposed on healthcare systems and the differing risk profiles of populations by location. The majority of exploration by location has been through decision modelling, which provides a framework whereby clinical effectiveness data (international or local) can be combined with local economic data that reflect the practice pattern of the location concerned. Incremental cost per life-year is the most frequently reported summary measure of cost-effectiveness in the studies looking at variability in results by location. The majority of the studies that have explored variability have compared European countries, with the UK and Germany most frequently cited, perhaps reflecting the diversity of healthcare systems in Europe.

Empirical analyses of geographical variability alongside studies using patient-level data

Studies looking at international variation

A number of studies have explored geographical variation in studies using patient-level data, usually as part of randomised trials. Most of these studies have considered the implications of changing unit costs from those used in the primary (i.e. base-case) analysis to those relevant to other locations. Other studies have considered in more detail the variation in resource use. These studies mostly employ location-specific unit costs to value resource use data collected across locations, or examine subgroups of patients defined in terms of the location in which they were treated.

Most of the studies identified examine **international** variation in costs and cost-effectiveness. Bennett and colleagues[56] investigated the resource use associated with the use of haematopoietic growth factors as adjunct therapy for autologous bone marrow transplantation for hospitals in Paris and New York. Although the clinical results were similar for both centres in that the use of granulocyte macrophage colony-stimulating factor (GM-CSF) was linked with a decrease in the duration of severe neutropenia, the economic results differed markedly due to differences in experience with the new therapy, practice patterns, institutional cost structures and healthcare systems. Significant savings relating to fewer days in hospital and fewer laboratory tests and radiographs for GM-CSF patients were noted for the New York hospital, but were not identified at the Parisian hospital. A random effects model was used to account for institutional variation in overall clinical results.

TABLE 6 *Characteristics of the studies looking at variability in results by location*

Factor	References
Clinical areas represented	
Bacterial infections and mycoses	83, 91–94, 97, 98
Cardiovascular diseases	34, 78, 79, 84, 87, 99, 100, 107
Digestive system diseases	31, 81, 115
Haemic and lymphatic diseases	89
Mental disorders	103
Musculoskeletal diseases	90
Neoplasms	56, 82, 95, 96
Nutritional and metabolic diseases	101, 102
Otorhinolaryngological diseases	105
Respiratory tract diseases	88, 106, 116
Virus diseases	80, 104
Methodological framework	
Clinical trial	34, 56, 78–84, 88, 100
Observational	87, 88
Economic model	31, 88–107, 115, 116
Cost-effectiveness summary measure	
Cost per life-year	78, 96–101, 104
Cost per complete cure	93
Cost per disease-free day/month	91, 94, 102
Cost per disease progression avoided	80
Cost per QALY	89, 106
Cost per symptom-free day	92, 103, 105
Cost per systemic therapy-free day	83
Cost per extra responder	95
Cost analyses	31, 34, 56, 81, 82, 84, 87, 88, 90, 115, 116
Locations	
A = Austria	84, 91, 115
AU = Australia	78, 80, 94, 115
B = Belgium	31, 78, 91, 95, 100, 105, 106
CA = Canada	34, 80, 83, 88, 91, 92, 100, 101
CH = Switzerland	91, 94, 102–104
D = Germany	78, 81, 82, 84, 89, 91–94, 97, 99, 101, 103, 105, 115, 116
DK = Denmark	104
E = Spain	78, 81, 83, 89, 91, 92, 95–97, 101, 103, 106
F = France	31, 56, 78, 81, 84, 89–92, 95–97, 101, 102, 104–106, 116
GR = Greece	91
I = Iceland	93
IR = Ireland	81
IT = Italy	78, 81, 82, 89–93, 97, 99, 101, 103, 104, 115, 116
LI = Lithuania	84
MA = Malawi	98
MO = Mozambique	98
NL = The Netherlands	91, 93, 95–97, 103, 105, 115
NO = Norway	78, 105
NZ = New Zealand	78
P = Portugal	78, 84, 91
PO = Poland	103
SA = South Africa	80, 100
SC = Scotland	106
SF = Finland	91, 93
SW = Sweden	78, 100, 102, 103, 106, 115
TA = Tanzania	98
TU = Turkey	83
UK = United Kingdom	31, 78, 81, 83, 84, 89–93, 96, 97, 99, 101–104, 115, 116
US = United States	31, 34, 56, 78, 79, 87, 99, 103, 115, 116
V = Venezuela	103
EU = European Union	34, 79, 80
Type of variation	
Between location	31, 34, 56, 78–84, 87–106, 115, 116
Over time	107

In another multinational trial, Schulman and colleagues[34,77] attempted to quantify resource utilisation and quality of life in a population with severe congestive heart failure and to explain the variation among patients receiving epoprostenol therapy versus best usual care alone in 14 countries. The study concluded that the validity of estimates of the costs and their differences may be limited, because they were determined within the environment of a Phase III trial and because the application of a single set of unit cost data would not demonstrate the true variation in unit cost estimates between locations. Therefore, it was recommended that the cost data be adapted to make them useful to different audiences.

Jonsson and colleagues[78] assessed the cost-effectiveness of treating coronary heart disease patients with simvastatin 20–40 mg once daily, using survival and cost data gathered prospectively in the Scandinavian Simvastatin Survival Study (4S). Incremental cost-effectiveness ratios were derived by combining national diagnosis related group-based hospitalisation and medication costs with the simvastatin and hospitalisation data gathered during 4S. Cost-effectiveness ratios derived by applying international costs to 4S data were similar across countries examined, ranging from £4137 per life-year saved in France to £8824 in New Zealand. Although data from 4S may be representative of other countries with healthcare systems similar to those Scandinavian countries in the study, the simple extrapolation of cost-effectiveness estimates to non-trial countries based on cost data alone may have limitations when applied to countries with markedly different clinical practice than in the 4S countries.

Holmes and colleagues assessed the association between use of resources and clinical outcome for patients with cardiogenic shock in the USA and other countries.[79] Resource use relating to interventions were compared from the GUSTO-I trial (1891 patients treated in the USA and 1081 treated in other countries). Diagnostic and therapeutic procedures were used more aggressively in the USA than in other countries – for example, in the USA, 483 (26%) of patients underwent percutaneous transluminal coronary angioplasty (PTCA) compared with 82 (8%) in other countries. Patients who underwent revascularisation had better survival in all countries, and adjusted 30-day mortality was significantly lower among patients treated in the USA than among those treated elsewhere (50 versus 66%, $p < 0.001$), with the difference in mortality remaining at 1 year. It was concluded that the lower mortality in the USA was probably a result of greater use of invasive diagnostic and therapeutic interventions.

Lacey and colleagues assessed the economic impact of adding lamivudine to zidivudine-containing antiretroviral regimens in patients with HIV infection.[80] Healthcare resource utilisation data (hospitalisation, unscheduled outpatient visits, medication and adverse events) were collected in a 1-year trial (CAESAR study) containing HIV patients from Canada, Australia, Europe and South Africa. A Cochran Mantel Haenzel test was used to control for country effects on treatment comparisons. To test for regional differences in resource use, a Mantel Haenzel χ^2 test was used to compare the rate of hospitalisations for HIV-related illnesses across all four participating regions. Resources were valued using UK and German unit costs. No statistically significant differences between countries in rates of hospitalisation for HIV were found, which the authors used to support the case for pooling the data. Reductions in resource use were associated with the clinical effectiveness of adding lamivudine to zidovudine-containing regimens.

Stalhammar and colleagues used data from a multinational trial to assess the cost-effectiveness of omeprazole and ranitidine when used as initial therapy in an intermittent treatment strategy for management of patients with symptomatic gastro-oesophageal reflux disease.[81] Trial-wide effectiveness data were combined with (1) a base-case analysis using trial-wide resources and country-specific prices and (2) country-specific resources and country-specific prices in the sensitivity analysis. Estimated mean direct medical costs were found to be lower for both dosages of omeprazole than for ranitidine in all countries except Germany, although none of these differences was statistically significant. Using country-specific data led to substantial changes in estimates relating to some of the countries: in France and Ireland, ranitidine became the least costly alternative, and in Italy and Spain, omeprazole 20 mg became the least costly alternative.

Mapelli and colleagues examined the economic consequences of the use of lenograstin as an adjunct to chemotherapy in women with inflammatory breast cancer.[82] Resource use data were collected in a trial involving 10 French hospitals. Resources were valued using corresponding German or Italian prices or costs for the same items. By reducing infection-related

morbidity associated with a high-dose chemotherapy regimen, lenograstin decreased treatment costs by DM 1794 and ItL 1.2 million, excluding cost of the lenograstin itself. As the lenograstin group reported fewer chemotherapy delays and hence benefited from more chemotherapy days, the chemotherapy treatment costs were DM 1519 and ItL 0.9 million higher than for the placebo group. Assuming costs of chemotherapy were the same for both groups, cost saving in the lenograstin group would be 30% in Germany and 34% in Italy. The underlying assumption was that patients treated in French centres would have been treated in a similar way in German and Italian hospitals.

Alongside a non-blind multinational clinical trial, Hakkaart-van Roijen and colleagues assessed the cost-effectiveness of tapered versus abrupt discontinuation of a microemulsion formulation of cyclosporin in patients with chronic plaque psoriasis in Canada, Spain, Turkey and the UK.[83] Tapered discontinuation was dominant (lower costs and more effective) in Spain, Turkey and UK and generated an incremental cost-effectiveness ratio of Can$1.4 per systemic therapy free day in Canada. In this study, relative price differences may have had an impact on cost-effectiveness as the price of cyclosporin was lower in Spain. Therefore, it was considered difficult to pool the cost data and the importance of reporting results of different cost components was emphasised.

In a conference abstract, Grieve tested the validity of the assumption that costs between countries are homogeneous in multinational economic evaluations.[84] For each stroke patient hospitalised during 1996 at centres in the UK, Austria, Germany, France, Poland and Lithuania, the duration of hospital stay, use of investigations and average medical time were recorded. Costs were measured using local prices and converted into dollars using purchasing power parities. It was found that the costs of stroke management varied across Europe because of differences in resource use and unit cost. Of the 576 patients in the study, the mean length of stay ranged from 12 days in France to 34 days in the UK. The average nursing time per day ranged from 113 minutes in Poland to 318 minutes in Austria. Also, unit costs were higher in western than in eastern Europe (e.g. a doctor cost US$34.21 per hour in France compared with US$1.32 in Lithuania). The lowest mean cost per case was in Lithuania (US$880, 95% CI 730 to 1030) and the highest in Austria (US$8336, 95% CI 6638 to 10,033). Therefore, the results of this study suggest that hospital costs measured in one European country are not generalisable to another and that the pooled data from multinational economic evaluations may not be applicable to each individual local setting.

Studies looking at intra-national variation

Some studies have considered geographical variability between centres within a single country. In one study, the importance of variation in unit costs was demonstrated in a non-randomised study comparing the costs and effects of routine mammography screening by single-view and two-view mammography.[85] The authors indicated that although the results were based on data from one screening centre and might be influenced by work patterns in that centre, the detailed reporting of the cost analysis undertaken would enable decision-makers in other centres to estimate how the results would differ for their situation. Other researchers have also arrived at similar conclusions in the analysis of trial data.[78,81,82,86]

Garson and colleagues assessed both the cost and the variation in pricing and practice for a chronic disease among six geographically representative centres in the USA.[87] For each congenital heart disease (CoHD) type classified according to physiological characteristics, the number of clinic visits, hospitalisations and years of medication use were estimated at each site. A 55% variation existed among the six centres in the charges for CoHD between birth and 21 years. The variation was unrelated to outcome (mortality) and approximately 50% could be accounted for by practice pattern. The greatest variation was in the use of complex clinic visits and cardiac catheterisation. The study showed that, by means of a multicentre approach, practice variation could be assessed and related to outcome.

Studies looking at variation between research and routine practice

Kennedy and colleagues explored the impact of being recruited as a subject for a randomised controlled trial (RCT) on the use and cost of asthma-related health services.[88] Because they controlled for geographical location, they were effectively exploring whether being treated in a trial had an effect on resource use and costs after controlling for geographical location. Resource utilisation (e.g. anti-asthma medications and ambulatory physician services) was compared and a logistic regression analysis was performed controlling for age group, asthma severity, year of data collection and geographic location. It was discovered that trial patients were more likely to use higher (400 μg or more) daily doses of

inhaled corticosteroids (ICSTs) than non-trial patients [odds ratio (OR) 3.1, 85% CI 1.6 to 6.2]; less likely to visit the emergency department (OR 0.4, 85% CI 0.2 to 0.8) and less likely to have two or more GP visits per year (OR 0.3, 85% CI 0.2 to 0.6). Log-transformed total asthma related costs did not differ between trial and non-trial patients. Certain categories of services (including ICSTs, emergency department physician visits and GP visits) differed in individuals taking ICSTs for their asthma, whether or not they had been enrolled in a clinical trial, but it was not possible to conclude that there was a difference in the total cost of asthma-related health services.

Empirical analyses of geographical variability in modelling studies

The use of decision models, in which costs and effectiveness estimates are synthesised from a range of sources, has proven to be a useful vehicle for exploring geographical variation in the cost and cost-effectiveness of healthcare interventions. These studies typically involve developing a core model (often based around the results of a specific trial) and then tailoring the model for individual locations (usually countries), usually based on location-specific resource use and/or unit costs. All the modelling studies identified considered variation between countries rather than between locations within a country.

Leese and colleagues estimated the costs and benefits of the use of recombinant human erythropoietin (epoetin) in the treatment of anaemia arising from chronic renal failure in France, Germany, Italy, Spain and the UK.[89] The incremental cost per QALY gained ranged from US$58,600 to US$349,000 depending on the assumptions made and the country. Each country's different approach to selecting patients for epoetin treatment and different healthcare financing arrangements led to varying treatment regimens and costs. The UK showed the largest average QALY gain, possibly reflecting the greater average severity of anaemia in the limited number of patients given epoetin in the UK.

Jansen and colleagues estimated the economic impact of the improved clinical tolerability of meloxicam compared with diclofenac in patients in France, Italy and the UK through the construction of country-specific models.[90] Probabilities of occurrence of adverse gastrointestinal events were established from RCTs and resources were determined from a medical database (UK), previously published literature (UK, France) and expert opinion (UK, Italy and France). Substantial differences in resource use were found to exist between countries for the same event. In France, length of hospitalisation for ulcer treatment was estimated by expert opinion to vary from 13.2 to 30.8 days compared with an average database derived value of 8.6 days in the UK. The most expensive unit cost related to hospitalisation associated with adverse gastrointestinal events, with an exponential increase occurring relative to the severity of the event. These rare hospitalisations had a strong impact on the average treatment cost in all countries for patients treated with diclofenac SR: 35, 10 and 13% for France, Italy and the UK, respectively. Potential cost savings resulting from the use of meloxicam from the viewpoint of the payer in each country were 32% (France), 5% (Italy) and 24% (UK) per patient per 30-day treatment. The French model was sensitive to variation in the probability of perforation, ulcer and bleeding and the Italian model was sensitive to probabilities of adverse events.

In another decision-modelling framework, Arikian and colleagues conducted a cost-effectiveness analysis comparing four oral therapies (griseofulvin, itraconazole, ketoconazole and terbinafine) for the treatment of onchomycosis of fingernails and toenails.[91] The analysis related to 13 countries: Austria, Belgium, Canada, Finland, France, Germany, Greece, Italy, The Netherlands, Portugal, Spain, Switzerland and the UK. Clinical data on success rates, relapse rates and side-effects were taken from a worldwide meta-analysis of randomised trials and combined with resource use devised around country-specific patient profiles and treatment algorithms. It was demonstrated that terbinafine was the most cost-effective therapy for both infections despite its higher acquisition cost in all countries in the study.

For the same disease, Einarson and colleagues constructed a decision analytic model to examine the expected costs and cost-effectiveness (in terms of costs per symptom-free day) of topical lacquers and oral agents in Canada, France, Germany, Italy, Spain and the UK.[92] Panels of experts determined clinical practice patterns, reimbursement practices and standard costs for healthcare resource items in each country. A meta-analysis of 33 studies, comprising 58 clinical arms, was used to determine cure rates. It was found that ciclopirox, as first-line therapy, had the lowest expected cost and lowest cost per symptom-free day, followed by amorolfine, terbinafine and itraconazole in all countries except the UK and Spain. The cost of treatment failure was much higher in the UK than for other countries.

Using data on complete cure rates from a randomised trial, Jansen and colleagues compared continuous terbinafine with intermittent itraconazole for fungal infection of the toenail in Finland, Germany, Iceland, Italy, The Netherlands and the UK.[93] Costs included medications, physician visits, laboratory tests, management of adverse events and management of relapse. Continuous terbinafine was dominant in Germany, Iceland, Italy, The Netherlands and the UK, whereas it had an incremental cost per additional complete cure in Finland. The study assumed that the only difference in medical management was the choice of antimycotic and the duration of treatment.

Shear and colleagues conducted a cost-effectiveness analysis of terbinafine versus ciclopirox, clotrimazole, ketoconazole and micronazole for the treatment of tinea infections in Austria, Germany and Switzerland.[94] The effectiveness data (efficacy rates, relapse rates) were based on a meta-analysis of clinical trials, and resource utilisation data (medications, physician visits, laboratory tests) were established from expert opinion. Terbinafine compared favourably in terms of cost-effectiveness with other therapies for all countries.

Annemans and colleagues assessed the impact of different management and different unit costs for healthcare resources on the cost-effectiveness of TAXCIS (paclitaxel + cisplatin) versus TENCIS (teniposide + cisplatin) for the treatment of advanced non-small cell lung cancer in The Netherlands, Belgium, France and Spain.[95] Interviews with clinicians, using a Delphi technique, and validation from patient chart analysis were employed to estimate resource use. A specific randomised trial provided effectiveness estimates. There were clear differences in practice and costs between the four countries, but this variation in medical practice did not influence the conclusion that TAXCIS was as cost-effective as TENCIS.

In a study by Van Ineveld and colleagues, economic modelling across locations was conducted to address the question of whether breast cancer screening programmes as in The Netherlands and the UK should be adopted by other EU countries.[96] A cost-effectiveness model for The Netherlands to assess the cost-effectiveness of a nationwide screening programme in women aged 50–70 years was adapted and populated with country-specific data on incidence, mortality, demography, screening organisation and prices for France, Spain and the UK. Cost-effectiveness varied by location: Spain (£9700 per life-year gained), France (£5800 per life-year gained), The Netherlands (£2120 per life-year gained) and the UK (£1800 per life-year gained), which the authors took as indicating that no uniform policy recommendations for breast cancer screening could be made across all countries in the EU.

Berger and colleagues compared the cost-effectiveness of paclitaxel–cisplatin (PC) combination therapy versus a standard cyclophosphamide–cisplatin (CC) regimen as first-line therapy in advanced ovarian cancer.[97] The effectiveness data were taken from a retrospective cohort study comparing cisplatin plus either cyclophosphamide or paclitaxel, and resource utilisation data (medication, hospitalisation, consultations, laboratory tests, investigations) were established by expert opinion. The PC combination compared favourably in cost-effectiveness in all countries with costs per life-year gained ranging from US$6396 in Spain to US$11,420 in Italy. The study demonstrated some differences between European countries concerning the proportion of total costs relating to chemotherapy and hospitalisation.

De Jonghe and colleagues compared the cost-effectiveness of short course and standard chemotherapy for pulmonary tuberculosis in Malawi, Mozambique and Tanzania.[98] Short-course chemotherapy was preferable to standard 12-month chemotherapy, and the cost-effectiveness ratios were remarkably similar across locations despite differences in the costs of food and labour. The authors commented that generalising the results to other developing countries may be questionable, however, because of differences in national incomes per capita where the cost of labour might be much higher in dollar terms.

Lorenzoni and colleagues evaluated the impact of differences in costs of thrombolytics on their cost-effectiveness in Germany, Italy, The UK and the USA.[99] The analysis was based on the crude costs of streptokinase and recombinant tissue plasminogen activator and 30-day mortality rates from a randomised trial. Even assuming common effectiveness between countries, the cost-effectiveness ratios varied widely by location: from US$112,344 per life-year saved in the UK to US$221,053 per life-year saved in the USA. The authors noted that the study's assumption of common effectiveness was called into question

because the clinical efficacy differed between European and American patients. This suggested that the variation they observed in cost-effectiveness may have been an underestimate of what would have emerged if variability in effectiveness had been incorporated into their analysis.

The West of Scotland Coronary Pravastatin Study (WOSCOPS) was conducted in Scotland, a country with a high incidence of cardiovascular disease. Caro and colleagues used location-specific information on the prevalence and clustering of risk factors to estimate baseline risks in Belgium, South Africa and Sweden.[100] The relative risk reduction from the trial was applied across all populations and combined with local costs in a model to estimate country-specific cost-effectiveness ratios: US$14,773 per life-year saved in Belgium, US$8150 per life-year saved in Sweden and $10,999 per life-year saved in South Africa. The authors concluded that the cost-effectiveness results from WOSCOPS were valid for other populations. Using similar methods, Grover and colleagues compared the cost-effectiveness of simvastatin for diabetic patients with cardiovascular disease (CVD) and those without in Canada, France, Germany, Italy, Spain and the UK.[101] Primary prevention among diabetic patients was as cost-effective as secondary prevention among CVD patients in all countries.

Ghatnekar and colleagues estimated the cost-effectiveness of treating diabetic foot ulcers with becaplermin plus good wound care (GWC) compared with GWC alone in a variety of European healthcare settings using a decision model with a 12-month time horizon.[102] Baseline probabilities were taken from a prospective study of 183 patients and becaplermin efficacy was based on 20-week healing rates in a meta-analysis of clinical trials with 449 patients. Country-specific economic data, established through expert opinion, were integrated into the model to generate cost-effectiveness ratios. Becaplermin was cost-saving in Sweden, Switzerland and the UK but had an incremental cost per ulcer-free month in France (US$142 per ulcer-free month). Substantial differences in resource costs existed, partly due to differences in unit costs and substitution effects between inpatient and outpatient settings.

A multinational decision analytic model was developed by Casciano and colleagues to examine the treatment of major depressive disorder in 10 European and American countries.[103] The effectiveness data (inpatient efficacy rates, outpatient efficacy rates, dropout rates due to lack of efficacy/adverse events) were taken from a meta-analysis of clinical trials, and resource utilisation was established by expert opinion. Venlafaxine XR was considered the most cost-effective in all countries except Poland on the basis of average cost-effectiveness ratios.

Using early economic data, Simpson and colleagues used a decision model to estimate the cost-effectiveness of adding zalcitabine to standard antiretroviral treatment for HIV patients in five countries: Switzerland, Denmark, France, Italy and the UK.[104] Standard treatment algorithms were developed using physician panels, and epidemiological data were adjusted to reflect the HIV/AIDS profiles in each country. Cost-effectiveness ratios for zalcitabine were similar in the five countries studied.

Pinto and colleagues compared the cost-effectiveness of emedastine, a new antihistamine, with levocabastine for the treatment of acute allergic conjunctivitis in Belgium, France, Germany, The Netherlands, Norway, Portugal and Sweden.[105] Effectiveness data (ocular redness, itching, days without symptoms and clinical failure) were derived from a clinical trial comparing emedastine 0.05% and levocabastine 0.05%, both twice daily for 42 days. An expert panel of ophthalmologists and GPs was used to establish the cost of first-line treatment failure including visits, drugs and laboratory tests. The cost of failure was lower for emedastine in all European countries and it was found to be economically dominant relative to levocabastine in Belgium, Germany, Portugal and Sweden, whereas in The Netherlands and Norway emedastine added to costs.

Ament and colleagues assessed the cost-effectiveness of pneumococcal vaccination in the prevention of invasive pneumococcal disease in Belgium, France, Scotland, Spain and Sweden.[106] The effectiveness data (incidence rates, mortality) were taken from a case–control study, and resource utilisation (hospitalisations, vaccines) was established from expert opinion. The cost-effectiveness of pneumococcal vaccination varied considerably across the five countries: Belgium (€25,907 per QALY), France (€19,182 per QALY), Scotland (€14,892 per QALY), Spain (€10,511 per QALY) and Sweden (€32,675 per QALY). The results were sensitive to mortality rates and the incidence of invasive disease, but it was suggested that country differences were not related to real variations in the magnitude of disease occurrence, but to differences in surveillance systems and case ascertainment.

Studies estimating variability over time

No examples of studies that have attempted to estimate variability over time were identified. However, in a conference abstract, Neumann identified a number of areas where variability over time may exist, including motor vehicle airbags, statins and implantable cardiac defibrillators (ICDs).[107]

Discussion

Virtually all the studies identified in the literature relate to issues concerned with variability in costs and cost-effectiveness **between locations**, rather than across time. A large number of studies were identified which considered factors which may generate variability in cost and/or cost-effectiveness between settings. Probably the most straightforward of these – and the most cited in the literature – are the unit costs associated with particular resources. Both between centres (within country) and between countries, it would be expected that variation may exist in the unit cost of resources such as hospital stay, clinical staff and outpatient attendances. It is likely that variation in some resources is more pronounced between rather than within countries, for example, pharmaceuticals and medical equipment. A number of factors cited in the literature as independent sources of variation in economic results may also partly determine differences in unit costs. This might include economies of scale, case mix and the skills and experience of relevant staff.

An interesting feature of the literature on sources of variation in economic studies is that some of the most frequently cited factors are as much associated with the measurement of **effectiveness** as with cost-effectiveness. This would include factors such as the artificial characteristics of centres undertaking research (e.g. randomised trials), patient case mix and clinical practice variation. However, it is probably reasonable to argue that issues of external validity and variability **in clinical effectiveness** by location are relatively rarely considered in the broader health services research literature. In considering some of the methodological implications for economic evaluation of variability by location, however, as much attention has to be paid to variation in effectiveness as to variation in costs.

Few studies were identified that considered explicitly the possible factors causing variability in the results of economic studies over time. However, there is a strong degree of consistency in the principles that can explain variability in costs and effectiveness between locations and across time. For example, just as patient case mix can vary from one hospital to another, so too it can vary over time within a hospital. This can be due to different clinical attitudes about the appropriate thresholds for intervention in a patient group, which can again vary by location and over time. It is also the case that factors such as unit costs, skills and experience and the healthcare system can vary over time in addition to by location.

A large number of studies have been identified which have explored variability based on both trials and decision models as the 'vehicles' for the evaluation. These include studies which have taken an existing trial or model focused on a specific location and incorporated a different set of unit costs. Some studies have collected resource use data in a number of centres and countries (often from clinical experts rather than direct observation) and then explored variation in those measures between locations. Not surprisingly given that a large number of factors can, in principle, lead to variation in economic variables, several authors have shown important differences in the volume and cost of resource use between locations. Relatively few studies have looked at variability in outcomes as assiduously as costs, reinforcing the view that most authors implicitly consider that clinical effectiveness measures are more exchangeable across locations than cost data. Most of the studies which have been reviewed in this chapter were standard economic evaluations which had sought to estimate results for a set of locations (usually different countries). Very few papers set out from the outset to measure variation between locations; and those that did (e.g. Holmes and colleagues' assessment of the association between use of resources and clinical outcome for patients with cardiogenic shock[79]) were more descriptive than evaluative.

Many healthcare technologies have been subject to a number of evaluations over time, but few studies have been identified which systematically compare these evaluations and seek to reach general conclusions about how costs and cost-effectiveness vary over time. Indeed, only one study[107] was identified which explicitly looked at this source of variability. This was only a conference abstract, but a fuller paper is expected in due course (P Neumann, Center for Risk Analysis, Harvard School of Public Health: personal communication, 2003) (since the review in this report was undertaken, a paper has been published which considers a range of issues regarding variability in cost-effectiveness over time[108]). This dearth of

literature may be due to the perception, on the part of researchers and policy-makers, that cost-effectiveness will not vary by time and location. However, given that many of the factors which this chapter has indicated will affect economics between locations, are also likely to vary over time (e.g. case mix, operator skill, prices), this seems difficult to accept. The second reason may be that, until recently, decision-makers only had the skills and capacity to consider the cost-effectiveness of new interventions once, if at all. However, that position is beginning to change as decision-makers acquire the additional resources to look iteratively at interventions. For example, NICE now reviews its guidance – and the cost-effectiveness analyses which underpin it – for technologies about every 3 years.

There are issues with respect to the systematic review methods. First, the systematic review presented here used two reviewers to identify relevant papers but only one to data extract. This was due to the large volume of literature relative to the resources available for the project. For a methodological project such as this, it is unlikely that only using one data extractor will introduce bias.

Second, searching on electronic databases for any methodological topic area, such as generalisability, is not an easy task. Comprehensive searching is even more problematic. There has been little research into conducting such search strategies and few indexing terms exist. For instance, there is no indexing term in MEDLINE or EMBASE that captures the concept of 'generalisability'. In addition to the absence of usable indexing terms, textword terms, such as 'generalisable' or 'transferable', frequently occur in the title and abstract in a different context to the one in which the project is interested. In addition, many papers mention generalisability in the negative sense; for example, "no generalisabilty was considered" or "this study may not be generalisable". This means that further irrelevant records will be retrieved and these records cannot safely be excluded by the search for fear of missing useful records.

A comprehensive or sensitive strategy which is usually required for a systematic search was, therefore, deemed impractical owing to the volume of irrelevant records that would be retrieved. A much more focused (precise), and hence pragmatic, search strategy was therefore developed. Although this strategy did not aim to identify all the research on this topic, it did attempt to find a large set of relevant papers. Handsearching in conjunction with the database searches proved invaluable. The handsearches identified an additional 31 papers (beyond the database searches), of which 21 were then found to have been indexed in at least one of the databases included in the searches.

The fact that these 21 papers were not found by the searches highlights the difficulty of achieving comprehensive (sensitive) literature searches for this topic while maintaining reasonable volumes of records. There are several explanations for why these papers were not found by the database searches. First, none of the 21 papers contained any indexing terms which would identify the topic of generalisability. Second, none of the 21 papers contained the term 'generalisable' or any synonym (such as external validity, transferable, extrapolated, portability) of this term. Four of the 21 papers contained no abstract, which reduces options for effective searching and retrieval. Seventeen papers (n = 21) did contain an abstract, but nine referred to geographical generalisability by either naming specific countries or citing a number of countries for comparison. Such wording would not have been retrieved by the search strategies because it was impractical to search for all possible combinations of countries that could be included. Three papers contained no indication that they had any relevance to generalisability in the bibliographic details or abstract (one was the 'unit costs of health and social care'). Their relevance was only realised by looking at the full paper. The remaining five papers contained vague references to generalisability that would not have been found by the search strategies.

Conclusions

This chapter has identified a large number of factors that are mentioned in the literature that might be expected to generate variation in the cost-effectiveness of health care interventions across locations. These include fairly obvious factors, such as the unit cost of particular resources, but extend to factors such as resource use, clinical practice and patient case mix. Many of these suggested sources of variation would be expected to apply to measures of clinical effectiveness in addition to cost-effectiveness. A number of papers were also identified which explored empirically variations in costs and cost-effectiveness between locations. Several papers have demonstrated differences in the volume and cost of resource use between locations, but few studies have looked at variability in outcomes. Only one conference abstract was located looking at variability in cost-effectiveness across time.

Chapter 4

A systematic review of methodological literature: methods used to assess variability and enhance generalisability

Chapter 3 describes the factors identified, conceptually and empirically, to explain likely variation in the cost-effectiveness of a given intervention between locations and over time. From this literature, it can be concluded that there is a reasonable basis to expect a series of interrelated parameters, relating to costs and clinical effects, to vary between locations and setting. This raises questions of method in economic studies regarding how variation in cost-effectiveness should be identified, quantified and interpreted. The review in this chapter summarises papers which have described and applied methods to assess variability and generalisability by location and time.

Basic results

Of the 109 references on the database, 52 studies consider methods that might be used in economic studies to explore variability and enhance generalisability relating to location. Those studies suggesting analytical methods are summarised in Appendix 2C; and those applying methods are detailed in Appendix 2D. The suggested methods fall into a series of categories as summarised in *Table 7*.

Assessing variability and generalisability in trial-based economic evaluation with respect to location

As described in Chapter 1, there is a well-established place for the prospective study, as a 'vehicle' for economic evaluation. That is, patient-level data from a single study are collected on resource use and outcomes and, when appropriately valued, these are used to estimate the cost-effectiveness of alternative interventions and programmes. These studies are frequently randomised trials,[117] but could also include non-experimental observational data.[118]

Within studies which collect data from a number of centres and perhaps in several countries, the opportunity arises to quantify variability between locations directly, and to assess whether this variability is sufficient to limit the generalisability of the study's conclusions – that is, the most cost-effective programme is not consistent between locations. In studies which collect data in one centre, the question is whether the results are generalisable outside that location, but data external to the study would have to be employed to assess this. Many studies collect a proportion of data (typically, resource use and clinical outcomes) from a number of centres/countries, but acquire valuation data (unit cost and, if appropriate, utilities) from only a subsample of locations (often only one). For these studies, scope exists to quantify variability relating to location directly, but external data may also need to be utilised to assess generalisability with respect to the valuation data. Several methods papers have been identified which consider questions associated with the generalisability of trial-based economic evaluation.

Methods relating to the design of trial-based economic evaluation

Authors have proposed a range of approaches that broadly relate to the design of randomised trials

TABLE 7 *Summary of the type of methods suggested to assess variability in economic evaluations*

Type	References
Analytical methods:	
Patient-level studies	21, 25, 33, 53, 77, 84, 98, 111, 112, 119–122, 124–129, 131–138, 141, 149, 157
Modelling studies	29, 31, 43, 65, 115, 116, 130, 139, 140, 142, 146, 158, 159
Critical appraisal methods	1, 26, 47, 114, 147, 148, 160

when these studies are used, in part, as a vehicle for economic evaluation. Methods identified in the literature relate to cost estimation, currency conversion, centre selection, randomisation, data collection and adjustments to bridge data from study to routine practice. These are considered in turn below.

Unit costs

The methods literature on the design and analysis of trial-based economic evaluation with respect to geographical variability have focused greatly on multinational trials – that is, how to interpret the results from a patient-level data collection exercise in a range of countries from the perspective of an individual jurisdiction. One aspect of the problem of interpretation relates to the unit costs used in the analyses. Multinational economic evaluations ideally require data on the unit costs of key resource items in various countries; this provides a necessary starting point to explore how costs and cost-effectiveness might vary between countries. However, obtaining unit cost data which have been consistently estimated across countries is difficult, partly because accounting practices vary greatly between countries. At a fundamental level, several authors have argued for the development of a standardised cost framework which includes the production of a checklist of common cost categories across countries, a standardised approach to service use data collection and agreed principles for estimating the unit costs of service use.[34,47,65]

It is common for unit costs from one country in the multinational trial to be used to proxy those in other countries and applied to pooled resource use data collected in all studies. For example, Jonsson and colleagues assumed that Swedish costs could be generalised to four other Scandinavian/Nordic countries in the 4S.[78] A similar approach is often used in multicentre trials within one country. However, some authors have argued for the importance of collecting country-specific or centre-specific unit costs, based partly on the view that unit costs are likely to vary systematically between countries for reasons described in Chapter 3.[53,77,84]

Furthermore, it has been argued that, in multinational trials, unit costs should be taken from all or a sample of centres within each country as applying unit costs from single centres in a multinational economic evaluation may reflect neither the average unit costs within a country nor the true variation in unit costs.[34] In a recent simulation exercise, it was confirmed that there is a significant difference in overall cost results between using unit costs averaged across centres and centre-specific costs to value all the resource use measured in the trial.[119]

Frequently, unit costs may not be available for some sites in a study. In a study of alternative management of subarachnoid haemorrhage, country-specific costing data were unavailable for eight out of 15 countries participating in the trial.[120] In this situation, a standardised costing methodology was developed in seven countries, and estimated unit costs were converted into a common currency using measures of purchasing power parity. A general linear model of cost was developed and the amount of variation in cost that could be explained by procedure and country was determined using a one-way analysis of variance. Despite the detailed approach adopted, there were still several resource items for which unit costs were unavailable for certain countries. For these items the analysts developed an index table, based on a market-basket approach. To estimate the cost of a given procedure, the market-basket estimation process required that cost information be available for at least one country. Where cost information was unavailable in all countries for a given procedure, costs were estimated using a method based on physician-work and practice-expense resource-based relative value units.

A conference abstract reported that the possible use of imputation methods to reduce the burden of unit cost data collection in a multi-national trial was explored.[121] It was found that the imputation error decreased as the number of types of hospitalisation and countries sampled increased, but that the rate of reduction in error shrank. Furthermore, the error was minimised by obtaining estimates for fewer types of hospitalisations from more countries. The authors concluded that these methods deserved greater attention when undertaking trial-based cost analyses.

Expressing costs in a common currency

Comparisons involving cost data from different countries also raise the issue of exchange rate conversions, although little attention has been given to the conversion of prices to a common base.[122] Simple published exchange rates are subject to fluctuations, do not reflect opportunity cost and may be misleading. Adjustments should be made using purchasing power parities (PPPs), and some economists argue that, where possible, PPPs should be specific to medical goods and services.[122,123] A further methodological concern is

that, although PPPs capture the relative cost of health technology purchases between countries, they fail to capture relative wealth (and thus ability to pay, which influences price) reflected in gross domestic price (GDP) PPPs. Another type of PPP – exchange rates and healthcare PPPs – has been tested for dialysis.[122] These are technology-specific and have been shown to cause least variation in costs between countries compared with other forms of conversion.[122]

Sullivan and colleagues developed an analysis plan for determining the cost-effectiveness of early intervention for asthma based on a multinational clinical trial. They stated their intention to use a regression analysis and to include a variable that would control for between-country cultural differences in the supply of healthcare, healthcare utilisation and patient behaviour.[124] Variables such as the number of hospitals or physicians *per capita* (as a proxy for health services availability) were considered. However, for reasons of data availability and to limit the number of variables in the analysis, the authors proposed the use of the annual gross national product (GNP) *per capita*, adjusted by PPP.

Identifying appropriate locations for trials

Although the randomised trial has become a very popular vehicle for economic evaluation, Chapter 3 has highlighted how the geographical location in which the study is undertaken may have a major impact on the results. As a result of this variation, several authors have discussed desirable features of trials undertaken for economic evaluation. In particular, there has been an emphasis on the need for careful selection of centres based on their economic or organisational characteristics and on the need for thought to be given to the definition of 'a representative centre' for trials.[53,125] To judge representativeness, Johnston and colleagues proposed that the average costs of trial centres are compared with the median values of non-trial centres.[125] To adjust the results of costs in trial centres, one-way sensitivity analysis was suggested to examine the impact on unit costs of each characteristic for which the trial centres were not representative. Alternatively, multiway sensitivity analysis can be used to explore the impact of several characteristics simultaneously.

Goeree and colleagues provided guidance on what might be taken into account when considering the generalisability of (hospital) unit costs.[53] The first consideration is the number of hospitals chosen for unit cost estimates. The larger the number of hospitals participating in the trial, the less representative or applicable are the costs from one hospital or a small number of hospitals. The second consideration is the method of sampling for selection of hospitals. This can be systematic or random. The selection of hospitals may be based on key hospital characteristics known to impact on unit costs such as hospital size, level of output, patient mix, teaching status, urban/rural location or extent of staff unionisation. The final consideration relates to the desired level of subgroup analysis by geographical area. Hospital cost variation may be as large within countries as it is between countries.

In attempting to adjust economic results based on data from trial centres in Argentina, Cuba, Saudi Arabia and Thailand to make them relevant to non-trial centres in South Africa, Gambia, Zimbabwe, Indonesia and Bangladesh, Mugford and colleagues collected additional data outside of the trial on local characteristics and features of the health service hypothesised to vary between locations including unit costs, morbidity and utilisation patterns.[126] These factors were considered because of country-specific governmental laws on the minimal healthcare for pregnant women, potential non-scheduled visits to the clinic that could affect outcomes and differing financing structures leading to different unit costs.

Dealing with trial-induced distortions to clinical practice

The issue of how to design trials in such a way as to make clinical practice in those trials more representative of routine practice has also been discussed in the literature. Coyle and Lee suggest the use of modelling, pragmatic trials and changes to data collection within RCTs to deal with the issue of protocol-driven costs – that is, resource use which is determined by the needs of the trial and which would not be expected in routine practice.[127] They argued that more realistic estimates of cost-effectiveness can be modelled based on efficacy results from trial data and data obtained outside the trial to provide a more precise estimate of economic impact free of protocol-induced effects. They use the example of a cost-effectiveness study of a particular group of pharmaceuticals, where efficacy results were taken from the trial and the costs of events were taken from non-trial sources. Also, study designs can be adapted to allow analysis of costs which have only occurred for protocol reasons in atypical centres to be modified. An alternative approach is to design data collection in such a way that it is limited to that which would be expected to occur in routine practice.

Dixon and colleagues link a number of approaches to provide more externally valid data on resource use and costs within the context of a randomised trial evaluating treatment for otitis media with effusion ('glue ear').[112] Scrutiny of the protocol identified four features of the trial that could potentially compromise the generalisability of the economic evaluation between locations such as the level of intensity of outpatient follow-up and assessment, the experience of the physicians (consultants) in undertaking operations, the level of intensity of preoperative assessments and the type of centres where recruitment took place. They suggest exploring simpler problems, such as unit cost bias produced by recruitment from selected centres, using one-way sensitivity analysis with routine cost data. More complex problems could be tackled with a series of observational designs to complement the data collected in the trial.

Others have suggested several unusual aspects of study design to increase generalisability.[128,129] These include allowing physicians to adjust dosages; recruiting a heterogeneous patient population with a variety of co-morbidities; administering treatment in an open-label fashion; following representative models of care; encouraging, but not enforcing, compliance with plans of treatment; following up irrespective of remaining on therapy; minimalising external monitoring; and stratifying costs according to possession of insurance coverage.

Methods relating to the analysis of trial-based economic evaluation

Most of the methods literature has focused on methods to assess and increase the generalisability of economic studies undertaken in parallel with RCTs. In recent years, an increasing number of these trials have been mounted on a multinational basis. The main motivation for undertaking multinational trials is that larger sample sizes can be assembled in a shorter period, thus allowing quicker, more precise estimates of treatment effect. In addition, because these trials enrol patients from a wide range of treatment settings and countries, they may increase the representativeness of the study sample and promote interest among clinical opinion leaders in several countries.

The multinational trial also provides a potentially valuable vehicle for economic evaluation. In particular, these studies provide an opportunity to explore variability in costs, effects and cost-effectiveness between countries, resulting from the sort of factors described earlier in this chapter.[77]

However, looking at estimates based on data pooled across countries may have limited applicability to any one country. Issues regarding appropriate methods to analyse multinational clinical trials are similar to those relating to multicentre trials conducted in one country: the value of accumulating more, and perhaps more clinically representative, patients needs to be balanced against the requirement to establish the cost-effectiveness of the interventions for a particular decision-maker whose interest may be focused on one centre or country.

The methodological problem of analysing multicentre/national trials

Although practice patterns and patient characteristics may differ across centres and countries in clinical trials, the starting point of most trial analyses is that the effectiveness of an intervention should not differ greatly across centres or countries, at least in relative terms.[31,130] Consequently, one should be able to pool the clinical data across centres and countries to assess the effect of treatment on clinical outcomes. The reasonableness of this assumption is sometimes assessed using formal hypothesis tests of heterogeneity, despite the fact that these are typically underpowered.[131] As noted above, the standard method of economic evaluation based on multinational trials is to apply a single set of unit costs (typically taken from one country) to all resource use data collected in the study (i.e. pooled across centres); these pooled costs are then related to pooled outcomes.

The assumption underlying these approaches is that resource use, in addition to effectiveness data, is perfectly exchangeable between countries and centres. However, Chapter 3 makes clear that this assumption may be a heroic one given variations between locations in factors which may impact on costs and outcomes. At the other extreme, multinational studies can assume that resource use data are not at all exchangeable between locations. Under this assumption, analyses for a particular jurisdiction would be based on the application of a single set of unit costs to resource use data taken from patients only recruited in one country in the trial. Although this may be feasible for large multinational trials, where a reasonably large number of patients are recruited in the country of interest, it will not be feasible for those countries which recruit relatively small numbers of patients. Related to this, some authors have concluded, through empirical investigation, that variations in healthcare costs between locations mean that they are fundamentally difficult to generalise, and a

more disaggregated approach to presenting economic evaluation data should be adopted rather than pooling inappropriately.[84,132]

It is likely that, in practice, both resource use and outcome data collected within trials where patients are recruited in a number of locations are partially exchangeable. That is, some components of costs and outcomes will be common between locations and others will be more specific to locations. Analytically, the problem is to disentangle these components. Several papers have been identified which have begun to explore these methods.

Estimating the relative impact of interventions

In describing the design of an economic evaluation alongside a multinational trial of a medicine used in the treatment for myocardial infarction, Jonsson and Weinstein argued that there would be differences between countries in the baseline level of resource utilisation (e.g. revascularisation procedures), since this had been seen in earlier studies.[33] In other words, they expected to see differences in 'standard practice' in the control arm of the trial. Therefore, they suggested calculating the pooled proportional difference in resource use across all countries, and then applying this proportional reduction to country-specific baseline resource use. This approach is analogous to that often used when estimating a clinical effect, where it is usually argued that the relative risk reduction from a given therapy is fairly constant across different study populations.[130] However, it remains an empirical matter whether the 'relative resource reduction' is constant across different healthcare systems, and this would need to be established in a given study before these methods were implemented.

Regression-based methods

Quantifying variability in cost-effectiveness

Several authors have suggested the use of regression methods to explore variability in costs and/or cost-effectiveness by location. Willke and colleagues carried out one of the few studies to suggest methods to assess variation in costs and outcomes together (i.e. in cost-effectiveness), rather than costs alone.[25] The analysis related to the use of tirilazad in the management of subarachnoid haemorrhage and used trial data from five countries. The approach assumed that the cost of an episode of care for an individual patient is determined by the treatment the patient receives, a series of exogenous variables (e.g. disease severity), the health outcome of the episode and country-by-treatment and country-by-outcome interaction terms. The model was specified in such a way as to 'decompose' the overall effect of a treatment on cost into the 'net' treatment effect (having removed the effect of outcome on cost) and the effect of outcome on cost weighted by the treatment effect on outcome. The country-by-treatment interaction term allowed for country-specific estimates of mean costs and outcome (mortality rate) for each treatment by country.

The authors demonstrated that, for five of the 11 countries in the trial, there were considerable differences in the average hospital costs and mortality rates (see *Table 8*). The results of the regression analyses showed that the full treatment effect of tirilazad on cost ranged from –US$4812 in country 5 to US$5845 in country 2, and that the statistical test for the country-by-treatment interactions tended towards significance (F-test, $p = 0.088$). The net treatment effect of tirilazad on cost ranged from –US$5041 in country 5 to US$4543 in country 2 – in this case, these differences were statistically significant ($p = 0.046$). The outcome effect on cost was negative in all countries, but also tended towards significant differences between countries. Finally, the country-specific effects of mortality were not statistically different from one another ($p = 0.64$).

TABLE 8 *Drummond and colleagues' checklist[1] to assess relevance of a published economic evaluation to a particular end-user*

Are the results valid?
Did the analysis provide a comparison of healthcare strategies?
Were the costs and outcomes properly measured and valued?
Was appropriate allowance made for uncertainties in the analysis?
Are estimates of costs and outcomes related to the baseline risk in the treatment population?
What were the results?
What were the incremental costs and outcomes of each strategy?
Do incremental costs and outcomes differ between subgroups?
How much does allowance for uncertainty change the results?
Will the results help in caring for my patients?
Are the treatment benefits worth the harms and costs?
Could my patients expect similar health outcomes?
Could I expect similar costs?

The regression estimates were then used to calculate incremental cost-effectiveness ratios for tirilazad treatment in three ways: (1) using own-country costs and mortality effects; (2) using own-country costs and the trial-wide mortality effect; and (3) using own-country prices, but trial-wide utilisation and mortality effect. The greatest variation in the country-specific ratios was in case (1), where country-specific resource utilisation, unit prices and outcome levels were all taken into account. In case (2), the ratios still varied considerably, although in three of the countries the ratios were similar, suggesting similar resource utilisation patterns. In case (3), when only prices were country-specific, there was much less variation in the cost-effectiveness ratios, suggesting that country-specific utilisation differences, controlling for outcome, clearly contribute to the variation in the cost-effectiveness ratios. The authors concluded that, although they found no significant country-specific differences in outcome (and hence outcome results were generalisable from place to place), simple transfer of trial-wide cost results to specific countries would have been inappropriate in their study.

Quantifying variability in costs

Other studies have used regression methods to assess variability in costs rather than cost-effectiveness. Rutten-Van Molken and colleagues used ordinary least-squares regression methods to look at variability in the cost of two alternative long-acting inhaled B_2-agonists between six European countries, based on data collected during a randomised trial.[133,134] Using country-specific unit costs, which were applied to the resource consumption of individual patients within each country, cost was regressed on a series of covariates including age, duration of illness, study drug and country. No country-by-treatment interactions were modelled. This regression method was then compared with another method where resource use was pooled, adjusted to reflect national medical practice better and country-specific unit costs were applied to the pooled volumes resulting in six different analyses (one per country).[134] The analyses of the pooled log transformed costs confirmed that there were no statistically significant differences in costs between treatments. However, significant differences existed between countries: Swiss patients had significantly higher costs than patients from other countries, whereas Italian patients had lower costs, which could be partly attributed to differences in definitions, costing methodology and uncertainty associated with data sources. It was concluded that it is difficult to pool the cost data from the six participating countries and the authors emphasised the importance of separating resource use from costs rather than simply reporting and analysing total costs.

Koopmanschap and colleagues suggest the exploration of the relationship between differences in medical practice and hospital medical consumption and costs through two approaches.[135] The first approach is aimed at finding differences in treatment patterns, given relatively homogeneous patient groups. For each of the main categories of medical consumption (e.g. hospital days, consultations, radiotherapy), a statistical analysis tests for country-specific differences in the amount of medical consumption, controlling for variables such as disease stage, treatment arm and age. If 'statistically significant and relevant' differences exist, these should be corrected for. For example, if in country X 20% fewer laboratory tests are carried out than in country Y (controlling for differences in patient characteristics), patients' laboratory consumption should be corrected downwards when estimating the costs as if all patients were treated in country X. The corrected amounts of medical consumption can then be multiplied with country-specific unit costs. An advantage of this method is that the results of the analysis are the amount of medical consumption in a particular country estimated as if all trial patients were treated in that country. A problem is that it is only feasible to make adjustments for relatively homogeneous medical services. The second approach aims to explain differences between countries more directly through explaining costs by age, disease state, treatment arm, country and relevant interaction terms. However, the authors recognise that it is not easy to decompose a cost difference into differences in the amount of medical consumption and the costs per unit of medical consumption. The study made no attempt to model variation in cost-effectiveness by location.

Another study, reported in a conference abstract,[136] also proposed adjusting observed resource quantity to a 'typical treatment' pattern of the target country. The authors also propose, first, substituting the unit prices of the target country; second, replacing selected treatment encounters with a functionally analogous service existing in target country; and third, sub-setting the original trial cohort to obtain outcomes in a patient population representative of the target country.

Rice and Jones suggest the use of multilevel regression modelling to detect centre-related

differences in randomised clinical trials since outcome (including cost-effectiveness) may be dependent on the characteristics of the centres themselves in addition to the randomised intervention.[21] Multilevel models can be used to analyse data that fall naturally into hierarchical structures consisting of multiple macro units (contexts) and multiple micro units within each macro unit. It can, therefore, be used to detect differences in centres and countries in a multinational trial.[137] This method is explored further in a case study in Chapter 7.

Other statistical methods

Papers have also explored the problem of distinguishing location-specific costs and effects using a range of statistical methods other than regression. Coyle and Drummond proposed the assessment of variation in costs between centres using one-way analysis of variance.[138] With these methods, patient-level variation in costs can be attributed to that between or within centres. For each treatment group in which differences in costs between treatment centres are statistically significant, analysis can be repeated to identify those components of costs for which there is significant variation between centres. For each of the cost components where variation between centres is evident, regression analysis can be conducted to assess the degree to which variation can be explained by various factors. This would allow identification of the causes of between-centre variation, thus assisting in determining the generalisability of results. The methods were applied to the CHART (continuous hyperfractionated accelerated radiotherapy trial) trial in cancer to identify the determinants of variation in radiotherapy and other hospital costs between patients and treatment centres. The results demonstrated that significant variation existed between centres only for CHART and specifically in relation to radiotherapy and hospital costs, implying that the overall cost results from individual centres may not be generalisable. Almost all the variation in radiotherapy costs could be explained by differences in unit costs between centres primarily due to differences in payment mechanisms for after-hours treatments and the annual use of equipment. For hospital costs, the provision of hostel accommodation and patients' characteristics were identified as important determinants of variation between treatments.

Cook and colleagues have proposed an approach for analysing multinational economic clinical trials based on testing for homogeneity in the data.[131] The general idea is to follow the approach used to analyse clinical data from such trials, where tests of homogeneity are typically performed before pooling the data. The authors' specific interest was in the possibility of a country-by-treatment interaction in the effect of treatment on the measure of effectiveness, cost or cost-effectiveness. The authors defined a qualitative (or crossover) interaction as occurring when the treatment effect is positive for the patients in some countries and negative for those in other countries. A quantitative (or non-crossover) interaction was defined as occurring when the magnitude, but not the direction, of treatment effect varies. If there is no evidence of treatment-by-centre interaction, the data can be pooled for analysis across centres/countries thus offering improved precision of the cost-effectiveness ratios for each country.

Assessing variability and generalisability with respect to location using decision models

A trial-based economic evaluation can only facilitate an analysis of variability in cost-effectiveness between locations if patient-level data are collected in all centres/countries of interest. If that is not the case, then it is necessary to employ data external to the trial to assess whether the results of a trial are generalisable to non-trial locations. Decision modelling is frequently used in economic evaluation for a range of reasons. One advantage of modelling is that it facilitates the synthesis of data from several sources.

Use of models to adjust trial results to reflect routine practice

One aspect of the potential role of decision modelling is to facilitate adjustment of the results of trials which are undertaken in centres and with patients which are known to be unrepresentative of routine clinical practice.

Some authors have suggested very general modelling approaches to adjust the results of trial. Rittenhouse proposes supplementing trial data with additional data to connect the artificial trial to the real world of clinical practice through modelling.[139] If data are not directly available, a Bayesian approach with priors based on expert opinion could be adopted. Using the example of a treatment option (HA-1A) for positive cases of Gram-negative bacteria (GNB) only in suspected sepsis, the above approach is used to correct for the highly selective patient groups in trials due to

strict inclusion and exclusion criteria. A major clinical trial in the field used strict inclusion criteria which enrolled a slightly higher proportion of GNB patients than would have been expected in the general population of sepsis patients. As a consequence, patients in the trial were more likely to benefit from HA-1A resulting in incremental costs per life saved of US$143,307–483,333. However, revising the trial results with estimates from 'real-world' diagnostic criteria derived from a study of house officers at a large university hospital, the incremental costs per life saved of HA-1A increased to US$211,538–707,143.

Baltussen and colleagues set the adjustment of trial results to reflect real-world practice within a three-step framework comprising successive assessment of internal validity, external validity and net impact at the system level.[114] Step 1 involves deriving internally valid results through collecting effectiveness parameters and patient-specific data on resource utilisation within a randomised trial. The second step involves adapting the results to the real world through taking into account context-specific factors such as specific physician, hospital and healthcare system characteristics. The methods which would be used in the second step include checklists, sensitivity analysis, use of observational data and use of pragmatic trials. The final step implies the assessment of costs and outcome changes associated with introducing the new intervention.

Substitution with location-specific data in a decision modelling framework

Decision models can be used to generate country-specific cost-effectiveness results.[1] This can take several forms, including the use of a model to extend the results of a specific trial-based economic evaluation to one or more locations (usually countries) which were not included in the trial. The importance of models in this context relates to the fact that, in certain countries such as the USA and Japan, regulatory authorities demand evidence of safety and efficacy from trials in their own populations. Therefore, trials are usually conducted in these localities, with and without economic data capture. However, although economic data may have been collected in a trial in one country (e.g. the USA), economic evaluations may also be required for other settings.

Decision models can also be used as a vehicle for economic evaluation more generally, where data on a range of parameters are taken from a number of sources, and this may involve estimating location-specific cost-effectiveness. This would include situations where a multicentre/multinational trial has established the effectiveness of particular interventions, but no trial-based economic evaluation was undertaken; a model can then be used to estimate cost-effectiveness in centres/countries within and outside the trial by synthesising trial data on treatment effects with other data necessary for cost-effectiveness analysis taken from other sources which can be partly location-specific and partly taken from a range of sources.

Models to extend results to non-trial locations

Bryan and Brown's study is an example of a study which used decision modelling as a framework to extrapolate the results of a study collecting patient-level data to a different (non-study) location.[30] The context of the study was the comparison of the costs and effects of routine mammography screening with a single view and two views of each breast. Although the analysis was based on data from one screening centre and the results might have been specific to its practice pattern, the cost analysis was reported in great detail so that other centres could estimate how results may differ for their situation. One issue with this sort of analysis is, as for trial analysis, if the production functions (i.e. the relationship between inputs and outputs) are fundamentally different between locations, then it is doubtful whether simple substitution can take into account differences between settings.[140]

Menzin used a decision model to extend the results of a US trial to patients in France, Germany, Italy and the UK.[116] Economic data were collected on patients in the trial comparing two different doses of rhDNase with placebo. In the trial, patients were treated for 24 weeks and the outcome measures included change in pulmonary function and incidence of respiratory tract infections (RTIs) requiring parenteral therapy. Resource use data on hospital admissions, inpatient days and days of oral and intravenous antibiotic therapy were also collected in the trial. In the US trial, the cost of treating RTIs over 24 weeks was US$1682 less among patients receiving rhDNase once daily than placebo, primarily due to reductions in the cost of hospitalisation.

The role of the model was to determine the economic impact of the therapy in France, Germany, Italy and the UK. Two key resource use parameters (rate of hospitalisation and length of stay) were identified in the early stages of the evaluation, which were then collected in all

countries. A simple approach was to price the US resource use observed in the trial in local currency, effectively ignoring differences in the production function between countries. This approach generated savings from rhDNase ranging from £434 (US$711) in the UK to FF 7010 (US$1064) in France. However, since the US trial-based resource estimates were potentially misleading, adjustments were made based on the variability in the two key parameters between the countries of interest. In the UK, no adjustments were considered necessary, but in Germany the length of stay was considered to be longer than in the USA (14.4 versus 12.3 days). In France and Italy, both the rate of hospitalisation and the length of stay were adjusted. The overall effect was to reduce slightly the estimates of savings in Italy (from US$908 to US$607) and in France (from US$1064 to US$850). In Germany, there was a very small increase in savings from treating RTI.

As emphasised earlier in this chapter, it is not only resource use and cost data that can vary by location; the same can apply to clinical data and this may also need to be reflected in models seeking to estimate location-specific cost-effectiveness. Several models in the literature have addressed this by assuming that the baseline risks for particular clinical events are location-specific whereas the relative treatment effect is more generalisable across locations. Caro and colleagues adopted this perspective and derived a generalised formula expressing cost-effectiveness in terms of elements that might be country-specific and those that can be assumed to be more general.[130] Within their framework, the relative treatment effect is generally transportable, but the baseline risk can be affected by differences in the distribution of risk factors in different populations. Similar methods have been proposed by others.[141]

The framework was used to adapt the clinical and economic findings from the WOSCOPS study, looking at the cost-effectiveness of primary prevention of coronary heart disease using prevastatin, from the location of the trial (the west of Scotland) to Belgium. The authors used local information on prevalence and clustering of risk factors in individual patients in a risk equation to estimate the reference (i.e. baseline) risk in Belgium; country-specific costs were also used. The relative risk reduction from WOSCOPS was then applied to the Belgium-specific data. The resulting absolute relative risk reduction in Belgium paralleled that observed in WOSCOPS (3.19% versus 3.48%). The cost-effectiveness ratio was €29,800 per life-year gained, similar to that in Scotland (€31,400). Therefore, it was concluded that the clinical and economic findings from WOSCOPS were generalisable to other populations with similar characteristics in terms of gender, age and cholesterol level.

Making general models location-specific

A number of examples exist in the literature of cost-effectiveness models which were developed to have general relevance to countries or centres, but where further work was undertaken to tailor the model to a specific location. In other words, there are models which initially seek to be generalisable across locations, but where country-specific analysis is subsequently considered appropriate. Discussing these issues in the international context, Haycox and colleagues outline three steps for making a general model specific to individual countries:[115] first, identification of variations in national cost-containment policies with their subsequent impact on clinical practice and resource utilisation; second, measurement of the impact of such variation on health and economic outcomes; and third, valuation of the extent to which such variations alter differential costs and effectiveness.

Drummond developed a model relating to the cost-effectiveness of misoprostol in the prevention of NSAID-associated gastric ulcer in patients with osteoarthritis who experienced abdominal pain.[31] The paper describes the use of a single structural model and core assumptions, which was then applied to the specific contexts of four healthcare systems: Belgium, France, the UK and the USA. All countries took core efficacy data from a US trial, but compliance with therapy, detection rates and consequent use of healthcare resources were allowed to vary by country. Cost comparisons between countries were made by adjusting to US dollars using purchasing power parities. The estimated net costs of 3 months of misoprostol prophylaxis showed surprisingly similar results, despite differences between the countries in clinical practice patterns and cost components. The main problems that this study encountered related to the availability and type of data found in each country; that is a common problem in modelling studies.

Alternative analytical approaches

Use of observational data for economic evaluation

The greater use of observational (non-experimental) data for economic evaluation has been suggested

as a possible way of increasing the generalisability of economic evaluations.[118] The attraction of this type of study is that it tends to be less selective in terms of patients included and hence has the potential to include a more representative sample of patients. However, in the absence of randomisation, comparisons of cost-effectiveness across non-randomised treatment groups raises concerns about selection bias. Selection bias arises as a result of the interaction of treatments and omitted or unobserved characteristics of patients that may influence treatment choice, but independently affect health outcomes.

A statistical approach which has been suggested to deal with selection bias is the 'decomposition technique'.[142] Originating in labour economics where it was used in the examination of wage differences between different population subgroups, the method was suggested as a way of adjusting trial efficacy data with observational 'real-world' data. The technique involves decomposing observed outcome effects into treatment and population components. The treatment effect is the difference in relative risk applied to the treatment group, whereas the population effect is an adjustment for differences in two study populations. Regression models are used to regress costs on sets of risk factors such as age, gender, race, disease group and logistic/linear regression used to estimate branch probabilities and payoffs within a decision-modelling framework. The practical potential of this technique lies in the fact that it enables policy-makers to identify the separate components of the outcome effect, which is crucial for decision-making. After decomposition, the treatment effect addresses the question of 'how much money will the new treatment save (cost) as compared to the existing treatment?' and the population effect considers the question 'how much money would have been spent by a potential treatment group receiving the new treatment compared with the control group?'. In other words, the treatment effect is the differential cost (benefit) resulting from the treatment itself, and the population effect is the differential cost (benefit) resulting from population differences, where the latter could be due to the difference between study to practice or different patient groups. This technique is also relevant in that it allows policy-makers to quantify how much the level of expenditure is likely to change following the adoption of the new intervention. Shih and Kauf[142] demonstrate the use of this technique in the context of a model of epoetin (EPO), a drug used in the management of patients with end-stage renal disease.

Methods have been described in the health economics literature for adjusting the results of observational studies for potential selection bias which go beyond the usual statistical adjustment of **known** potential confounders.[143,144] These methods involve the use of instrumental variables (IVs), which are variables (or 'instruments') that are 'predictors' of the particular treatment which is selected for a given patient, but which are not correlated with the outcome of interest. When this assumption is valid (i.e. when an appropriate instrument has been identified), the method mimics randomisation by identifying 'balanced' sources of variation in treatments which generate estimates of treatment effects which are not affected by selection bias. However, these methods generate rather different estimates from randomised trials because the groups being compared differ not in the treatments they receive but in the **likelihoods** of treatment. This results in the estimation of an incremental or 'marginal' treatment effect which relates only to the range of variation in treatment across the IV groups.

McClellan and Newhouse applied IV techniques to an economic evaluation study.[144] The study estimated the incremental mortality and costs of intensive procedures in the elderly. These interventions were cardiac catheterisation and subsequent revascularisation (angioplasty or bypass surgery), and the study used panel data relating to the US elderly population from 1987 to 1990. A clinical evaluation of the interventions, using the same methods, had been published earlier.[145] The particular instrument selected for the analysis was the distance that a patient lived from a hospital offering more intensive treatment, on the basis that this variable was a good predictor of treatment allocation but was assumed (and to some extent shown to be) independent of outcome. The method estimated the average effect of invasive management for all patients who are 'marginal' (that is, who undergo invasive treatment in the 'relatively near' group but not in the 'relatively far' group) given that the groups are balanced in observable characteristics and that there are no other treatment differences between the groups. The most conservative cost-effectiveness estimate for intensive technologies was at least US$40,000 per patient surviving to 1 year. In contrast, ordinary least-squares estimates, which adjust for observable patient differences between treatment groups but which do not allow for selection bias, generated a cost-effectiveness ratio of under US$30,000. This led the authors to suggest that much of the existing literature on medical technology assessment based

on observational data may involve substantially biased estimates. McClellan and Newhouse proposed that this method could be more relevant for policy analysis as it involves estimating the incremental average cost-effectiveness of **changes** in medical treatment patterns rather than the population average cost-effectiveness for an observable population. However, in practice, the validity of the IV approach relies on finding instruments which satisfy the key condition of being good predictors of treatment allocation but are not correlated with outcomes.[23]

Manning and Claxton discuss the more general use of IVs to increase the generalisability of clinical trials, for both clinical and economic evaluation.[143] They describe the well-understood problem that clinical trials usually do not include a random sample of patients, hospitals and clinicians from the population for whom the trial is supposed to the relevant. This can be partially addressed by using regression methods to explore interactions between treatment allocation and covariates relating to patients (e.g. age, co-morbidities) and providers, assuming that trial patients exhaust the full range of these variables. (This method is explored further in the case study in Chapter 6.) The problem is that this 'subgroup analysis' only relates to the **observed** characteristics of the trial sample and the wider population. The paper argues that, given that most patients in trials are selected for randomisation (either by themselves or by clinicians) rather than randomly drawn, there may be **unobservable** differences between the trial sample and the wider population. To the extent that there is an interaction between treatment allocation and these unobserved characteristics, the average treatment effect observed in the trial will be a biased estimate of that which would apply in the wider population. This is considered to be a manifestation of selection bias similar in its implications to that which is associated with observational studies. It is argued that this problem can only be addressed by modelling the selection of patients into trials using methods such as IVs.

Generalised cost-effectiveness analysis

Some other approaches to assessing variability and generalisability in economic evaluation have been proposed that do not strictly fall into trial-based or modelling frameworks. Murray and colleagues proposed a novel approach to economic evaluation that could make the results of economic evaluation more useful to individual settings.[29] Given that the proposed methods would inevitably involve the synthesis of information from a range of sources, a modelling framework would be needed for implementation. Their guidelines on 'generalised cost-effectiveness analysis (CEA)' propose the application of CEA to provide general information on the relative costs and health benefits of different health interventions in the absence of various local decision constraints. 'Generalised CEA' involves the evaluation of the costs and benefits of a set of related interventions with respect to the counterfactual of the null set of the related interventions – that is, comparing costs and benefits with a particular intervention and without any intervention (i.e. a 'do nothing' option). According to this approach, this provides the complete set of information for evaluating both independent and mutually exclusive options to identify the health maximising combination of interventions for any given budget.

The results of 'generalised CEA' are then presented in a single league table. For each set of mutually exclusive interventions, the intervention with the lowest cost-effectiveness ratio with respect to a 'do nothing' option should appear first in the league table. This is akin to selection on the basis of an average cost-effectiveness ratio. The second intervention from the set (if there are at least two) that appears in the league table is the one with the incremental cost-effectiveness ratio with respect to the intervention that already appeared in the table. The third intervention is the one with the lowest incremental cost-effectiveness ratio with respect to the second intervention, and so on. By analysing the costs and benefits of mutually exclusive interventions with respect to a notional 'do nothing' option, the results were considered likely to be more transferable from one population to another, although the authors admit that this will be confirmed only through experience. There are clear limits to the comparability across populations of the 'do nothing' option which will depend on the development of the health system and on epidemiological patterns. The authors concede that their methods would be a technical challenge since estimating the costs and benefits of 'do nothing' options for a group of related interventions will require the development of natural history models.

Critical appraisal methods

A fairly extensive literature exists on methods for critically appraising economic evaluations. As part of this, there has been some interest in the use of systematic overviews of economic evaluations as a way of assessing the consistency of results across different locations. Heyland and colleagues suggest the use of a checklist, shown in *Table 9*, which can

TABLE 9 *Mean variations in total cost and mortality rate by country for trial patients in Willke and colleagues' study*[25]

Country	Average hospital cost per patient (US$)	Mortality rate
Country 1	18,180	0.200
Country 2	14,476	0.206
Country 3	14,007	0.111
Country 4	19,561	0.257
Country 5	41,258	0.084
Test of equality of country means	$F = 66.6$, $p < 0.0005$	$\chi^2 = 23.6$, $p < 0.0005$

TABLE 10 *Heyland and colleagues' checklist*[26] *to assess the generalisability of economic evaluations*

Clinical generalisability
Are the patients described in the analysis similar to those patients you see in your own setting?
Systems generalisability
Are the patients described in the analysis similar to those patients you see in your own setting?
Is the intervention under study generalisable to your setting? (i.e. despite good clinical evidence, is such a programme available or likely to be available in your setting?)
Are the costing methods applicable to the healthcare system in which you work?
Is the unit price for drugs, physician fees, laboratory tests, etc., the same?
Is the mix of resources consumed the same?
Is the volume of patients, and therefore the average cost per patient, similar across systems?
Can you convert exchange rates across systems appropriately?
Are the outcomes measured appropriately to your setting?
Was a method to measure the outcomes compatible with the current methods utilised in your setting?
If a preference-based measure was used, is there evidence that the preferences of your patients are the same as those preferences used in the analysis?
Is the discount rate applicable to your setting?

be applied to published economic evaluations to assess how applicable the study is to the particular location of interest.[26] The authors suggest criteria at two levels – clinical generalisability and systems generalisability – and ask the reader whether the clinical and economic data in a study are relevant to their context. In similar vein, Drummond and colleagues propose a series of questions for the end-user to determine whether an economic analysis yields valid and important results, which can then be applied to his/her clinical setting.[1] These questions are shown in *Table 10*.

Several studies have used secondary reviews of existing economic studies to assess variability in results between locations. Jefferson and colleagues performed a systematic overview of epidemiological and economic variables for the prevention and treatment of influenza.[140,146] The authors attempted to construct a secondary economic model which took data on resource inputs from studies identified in the literature and estimated costs and cost-effectiveness from unit cost data specific to a particular location. Through pooling, ranges of resource estimates were constructed, to which standard costs were attached. These were then used in a series of evaluations to determine the costs and benefits under different scenarios while maintaining a fixed efficacy of the vaccine throughout.

In the context of the French healthcare system, Spath and colleagues propose a three-step methodological approach to evaluate the potential of studies for transfer.[147] First, they carried out a literature search of economic studies. Second, they conducted a critical appraisal of the studies based on four inclusion criteria: identification of perspective, two or more competing options, description of therapies and relevance of therapies to the French healthcare system. Third, they undertook an analysis of the eligibility for transfer based on indicators which included the settings in which the studies might be used, the transferability of health outcome data and the transferability of resource utilisation data. This framework was applied to economic studies relating to breast cancer treatment to assess their transferability to the French healthcare system. Of the 26 studies identified in the literature, 20 did not satisfy all four inclusion criteria. Six studies were appraised but none of these studies were

considered relevant to the French healthcare system. The main reason for this conclusion was that the cost data were not reported in a sufficiently transparent way.

A similar approach has also been proposed, in a conference abstract, which groups transferability factors by methodological context, health care system and population characteristics.[148] The main purpose of this checklist is to allow a quick determination of possible transfer problems, the identification of the most needed adjustments and 'knock-out' factors which may preclude transfer (e.g. great difference in disease incidence).

Davis proposes that there are two levels to the consideration of the transferability of a study: mechanistic and economic.[149] The mechanistic question is whether the same pathological mechanism of therapy would apply in other segments of the population. The economic question is whether the treatment would be cost-effective in populations other than those studied in the trial. Davis[149] argues that to be able to judge the extent to which the results of a clinical trial can be generalised requires additional information. This includes the evaluation of other evidence including basic science laboratory studies, animal studies, genetic studies, observational studies and other RCTs with similar settings or treatments.

Other checklists attempt to assess the extent to which a trial-based economic analysis represents the routine practice of a particular decision-maker or the degree to which sufficient information is reported for a judgement to be made. For example, Baltussen and colleagues suggest that five questions should be addressed when reporting the results of a trial-based economic analysis:[38] (1) is epidemiological information reported?; (2) are context-specific factors explicitly reported?; (3) is an indication of the impact of non-compliance given?; (4) has 'real world' sensitivity analysis been applied?; and (5) are future developments indicated?

Methods relating to time

No studies have been identified which directly consider methods to assess variability in cost or cost-effectiveness over time. A growing area of research is developing on methods for iterative technology assessment: the process of establishing the cost-effectiveness of interventions based on best available existing evidence and modelling techniques, and then using the same models to identify priorities for future research.[28,150] This includes a major new area of methodological research in economic evaluation – Bayesian analytical approaches.[12,151–153] However, these methods are much more general than assessing time-related variability in cost-effectiveness, so they have not been included here.

Discussion

As shown in Chapter 3, there are good reasons to think that the results of economic evaluation will vary by location. A key issue for this report is the implications that this has for economic evaluation methods in the future. A number of specific questions will need to be addressed in particular types of economic study.

How can generalisability to other locations be assessed in primary studies (e.g. trials) undertaken in single locations?

One way in which authors have sought to address this issue is in terms of the design and reporting of studies. Papers have been identified which urge more careful thought about the selection of the location for trials,[53,125] and more 'pragmatic' study design features such as allowing a more varied patient population.[128,129] The literature also considers methods for the critical review of economic studies, including the quality of reporting in studies to allow local decision-makers to assess the relevance to their context.[26]

Although the design and reporting of these single-location studies is a necessary part of assessing likely generalisability, it is unlikely to be sufficient. In many cases, there will be good reasons to think that factors such as patient mix, resource use or unit costs are different in locations other than the one chosen for the primary study. Analytical methods are then required to adapt the results of the primary study. In effect, these 'extension studies' are assessing how robust the results of the primary study are to variation in key aspects of cost and/or effect that might be expected in other locations. In the papers reviewed in this chapter, the decision model is the main means by which this sort of adaptation is undertaken. Usually, the focus is to take the results of a trial undertaken in one country and extrapolating to another – for example, Caro and colleagues' analysis to extrapolate from the WOSCOPS trial in Scotland to clinical practice in Belgium.[130] However, the principles also apply to extrapolation between centres within a single country.

Although most of the extrapolation models reviewed in this chapter focus on changes to the cost side of the evaluation, it is clear that there is a range of reasons why the effect side may also need to be adapted from one setting to another. The methods issues related to generalising and adapting clinical effectiveness have been considered elsewhere,[19] and are not the focus of this report. However, these principles are also highly relevant to CEA. To the extent that these methods have emerged in the literature reviewed here, they involve the view that baseline event rates in studies tend to be location-specific, whereas the relative treatment effect is more exchangeable across subgroups and locations.[130] In this case, the decision model becomes the framework to synthesise the relative treatment effect from the trial with location-specific baseline risks. It is important to note, of course, that the assumption about the exchangeability of relative treatment effects should, as far as possible, be tested.[19]

How can generalisability across locations be assessed in primary studies (e.g. trials) undertaken in several locations?

The multinational trial has emerged as an important area of clinical evaluation in recent years. This is probably due to the fact that such studies provide an efficient means of recruiting large numbers of patients for regulatory trials for pharmaceutical companies, and for the 'large and simple' trials where the emphasis is on the need for sufficient power to detect small but important treatment effects.[154] An important premise of such trials is that the nature of the technology (invariably a pharmaceutical) is that its treatment effect will be fairly generalisable across locations. There has also been a growth of multicentre trials within single countries, again in an attempt to recruit sufficient numbers of patients rapidly. These multicentre studies also hope to be more generalisable, but this should be demonstrated analytically. Given that these trials are frequently vehicles for economic evaluation, this chapter has highlighted a growing methodological literature on appropriate analytical methods in this area.

The starting point of the papers reviewed here is that, for the range of reasons which have been discussed, variables within an economic evaluation are likely to vary between locations. However, a further factor is that basic economic theory would suggest that these local variables are related through the production function. For example, if the labour cost of a particular medical professional is relatively high in a specific location (perhaps owing to a shortage of that group), a decision may well be taken locally to substitute other inputs into healthcare delivery (e.g. appropriately skilled nurses). In turn, this may or may not impact on health outcomes.

The challenge, in terms of the methods used to undertake economic evaluation alongside multilocation trials, is how to reflect this inter-dependence between costs, resource use and outcomes, and the likely variation between locations. A standard analytical approach is, essentially, to ignore this inter-dependence by applying unit costs from one or a few centres/countries to pooled resource use, and to relate costs to pooled outcome data. For example, in a trial comparing high- and low-dose ACE inhibitors Sculpher and colleagues used UK unit costs to value the resource use of all patients, regardless of the system within which they were treated, and differential costs were related to differential life-years which were again pooled across all trial patients.[155] An implicit assumption of these analyses is that resource use and outcomes are exchangeable across locations. An alternative assumption is that resource use and outcomes are not exchangeable across locations, and that it is only possible to estimate the cost-effectiveness of an intervention in a given location by collecting relevant data there. This approach suffers from the problem that few multilocation trials will recruit sufficient patients to allow reasonably precise estimates of location-specific cost-effectiveness. The review of applied multilocation trials in Chapter 5 will shed more light on the economic evaluation methods being used alongside these studies.

An emerging methods literature has been reviewed here that is seeking appropriate forms of analysis which accept that some components of resource use and outcomes are exchangeable across locations whereas some are not. Some authors have explored the value of regression analysis, but confined their analysis to an exploration of variation in cost (rather than cost-effectiveness) between locations.[133–135] Two key studies have been located looking at methods of assessing variability between centres in cost-effectiveness. Cook and colleagues[131] look to apply the tests of heterogeneity to economic data in a similar way to how they are used with clinical data. Essentially, these methods seek to establish whether it is reasonable to pool data across locations. If the statistical tests show evidence of variation between locations, pooling should be avoided. The dataset used by Cook and colleagues to demonstrate their methods showed no apparent

indication of heterogeneity. The limitation with these methods is that they rely on a dichotomous decision about pooling based on a statistical test which needs a decision to be made regarding a 'reasonable level of heterogeneity'.

Willke and colleagues use a more complex approach to model location-specific cost-effectiveness.[25] Their regression methods are effectively based on decomposing the cost-effectiveness of an intervention into a series of separate effects, each of which can be adjusted for the location in which the patient was treated. Recently, methods have been developed to place cost-effectiveness analysis into a general regression framework.[156] This opens up a range of possible analytical techniques to assess variation in cost-effectiveness across locations. This theme is explored in detail in a case study in Chapter 7 which uses multilevel regression modelling as a means of estimating location-specific cost-effectiveness.

Finally, there will always be a need to extrapolate cost-effectiveness results from multilocation trials to one or more locations which did not recruit patients. Future development of regression methods may allow this to be undertaken more systematically using covariates to characterise centres. One potentially powerful analytical framework is the use of selection models and, in particular, instrumental variables. This has been described in the health economics literature[143] but not yet widely applied. It can be used to model how patients are allocated to different treatments in observational studies to increase the internal validity of those studies. It can also be used to increase the external validity of trials by formally modelling how patients (and providers) are selected for those studies.

How can generalisability and variability be reflected in economic evaluation decision models?

The key role of the decision model as described in this chapter has been as a way to adjust the cost-effectiveness estimates in primary studies, such as randomised trials, to locations which did not recruit to those studies. However, for a range of reasons, the decision model will increasingly be used as a general framework for economic evaluation.[12] These reasons partly relate to the limitations of trials for cost-effectiveness analysis (e.g. inappropriate or partial comparisons, use of intermediate end-points, insufficient measurement or short follow-up) and partly to the dearth of trials in many areas. Hence the decision model should be seen as the main vehicle for cost-effectiveness by synthesising data from a range of alternative sources. However, the familiar issues relating to variability between locations in cost and cost-effectiveness will emerge in this process of evidence synthesis. When deciding which data to use to populate a model, the extent to which they are directly applicable to the particular decision-maker who is the target of the analysis needs to be considered. As discussed above, adjustments will be necessary if data are not considered exchangeable between locations. These issues are explored further in the review of applied modelling studies in Chapter 7 and the modelling case study in Chapter 8.

Variability in cost-effectiveness over time

This systematic review failed to identify a major methods literature on variability in cost-effectiveness over time. One conference abstract had sought to quantify the extent to which the cost-effectiveness of motor vehicle airbags, statins and ICDs had varied over time,[107] but this area has not been widely researched. The reasons for this are likely to be similar to those considered in Chapter 3.

There is, however, a growing literature on the use of iterative methods to evaluate healthcare technologies,[28,150] which was considered outside the scope of this review. These methods are based on the premise that the cost-effectiveness of a technology will vary over time as new research data emerge about its use. More fundamentally, formal economic analysis and statistical decision theory are used to identify the priorities for additional research.[151–153] A dynamic process emerges whereby a common decision analytic framework is used to inform decision-making about the use of particular technologies based on existing information and about the needs for future research and its optimal design. As new information emerges, this Bayesian framework is used to update the prior estimates of cost-effectiveness.

Conclusions

Methods to assess generalisability and variability in economic evaluation studies have been discussed extensively in the literature. These relate to both trial-based and modelling studies. For the former, regression-based methods are likely to offer a systematic approach to quantifying variability in results between locations in multicentre/country trials. Decision analytic models are likely to retain an important role in adapting the results of cost-effectiveness studies between locations.

Chapter 5

A systematic review of economic evaluations undertaken alongside multicentre randomised controlled trials

Most of the methods literature on generalisability has focused on issues relating to economic studies undertaken in parallel with RCTs.[46,109,117] RCTs are often designed to randomise patients recruited from a number of different locations, where these can be represented by several centres within a country and/or by more than one country. For both clinical and economic evaluation, these multicentre or multinational trials often offer several potential advantages over single-centre trials. These include the opportunity to ensure speedy recruitment of the target sample size and a means of meeting the needs of regulatory agencies in different jurisdictions.

In addition, the use of this sort of trial design also has the potential to increase the generalisability of estimates of effectiveness and cost-effectiveness. This can be achieved in two ways. First, by recruiting patients from a range of locations, there is an opportunity for those patients recruited to the trial to be more representative of those presenting in clinical practice. Second, it is possible to assess how variable cost and effect results are within and between settings, and hence to establish whether the overall results of the study are generalisable across locations. In the context of multi-national trials, the second of these factors is particularly important given the proliferation of healthcare systems now requiring economic evidence to decide whether a particular health technology should be reimbursed.[5] This is because, in principle, a single multinational trial can provide necessary cost-effectiveness evidence for a number of jurisdictions.

The potential value of multilocation trials as a way of increasing generalisability in evaluation is relevant to both clinical and economic studies. However, for the reasons described in Chapter 3, numerous factors suggest that cost-effectiveness estimates are particularly sensitive to the location from which the data were drawn. As described in Chapter 4, a number of papers have considered the question of how to analyse economic evaluations alongside international trials.[25,77,161] These methodological issues also apply to the analysis of multicentre trials in a single country.

It is not clear, however, whether a recognition of the potential value of multilocation trials for generating data on clinical and economic variables in representative samples of patients or recently suggested methods for assessing variability in cost-effectiveness between different centres and countries has yet permeated into applied economic studies. To address this question, this chapter reports the results of a systematic literature review of published economic evaluations alongside multinational or multicentre RCTs. The main emphasis is a critical appraisal of the way in which published studies allow the reader to appreciate the extent of variability of the study's results across centres and/or countries and its implications on the generalisability of the study's findings. To achieve the above objective, the review addresses the following three main questions:

- Are economic studies reporting their results in a way that allows decision-makers to assess the study relevance and generalisability to their own location?
- What methods (if any) are being used to explore the variability of results between those locations participating in the trial?
- What methods (if any) are being used to assess the generalisability of results to locations not participating in the trial?

Several authors have contributed to the methodological discussion regarding what is to be considered good-quality reporting in economic evaluation studies conducted alongside RCTs.[25,162,163] Nevertheless, the specific issue of what to report to enable decision-makers to evaluate the relevance and generalisability of the study results to their own location has received less attention.

Methods

Inclusion criteria

The review aimed to identify full economic evaluations which were undertaken alongside multicentre or multinational RCTs and focused on the relevant literature published, in any language, during the period 1994–2000 inclusive. To minimise the risk of missing potentially relevant contributions, the inclusion criteria for the present review have been left intentionally broad. Furthermore, in view of the fact that a large number of methodological contributions in this area of research have been published in 1994 and thereafter (see Chapters 3 and 4), it was envisaged that the decision to limit the search to studies not older than 1994 would not undermine the relevance of the study findings. Furthermore, the inclusion of older studies in the sample would have biased the findings, artificially suggesting a poorer level of reporting quality.

Search strategy

The only source of the literature searched was the NHS Economic Evaluation Database (NHS EED).[164] The NHS EED is a database of detailed structured abstracts of economic evaluations written by a team of health economists, which contains comments on the reliability of the study. The range of record types on the NHS EED include abstracts of healthcare-related cost–benefit analyses, cost–utility analyses (CUAs) and CEAs (including cost minimisation and cost consequences analyses); brief records of costing studies; brief records of articles on the methodology of economic evaluations; brief records of reviews of economic evaluations; and short abstracts of records from the Department of Health's Register of Cost Effectiveness Studies. The database is freely accessible from the web page of the Centre for Reviews and Dissemination (CRD), University of York (http://agatha.york.ac.uk/nhsdhp.htm), and a detailed guide to searching the NHS EED can be downloaded from the URL: http://agatha.york.ac.uk/EEDSEARCHING.doc.

The NHS EED includes economic evaluations identified from more general databases such as MEDLINE, CINAHL and EMBASE and also through handsearching journals and working papers. Its inclusion criteria ensure that only economic evaluations that meet the project's definitions of a cost–benefit analysis, cost-effectiveness analysis or CUA[164] are selected for abstracting.

BOX 1 *Search Strategies used for the NHS EED and CAIRS software*

centres or centers or multi$ or countr$ or nation$ or multinational or international) and (randomised or randomized or RCT)
prospective$ and not retrospective$

Although an alternative database of economic evaluation studies exists, namely the Health Economic Evaluations Database (HEED), the NHS EED was preferred to HEED because the former searches more sources, its abstracts are more detailed and it has better search and export facilities. Although HEED goes back further than NHS EED, NHS EED was preferred also because its categorisations permit economic studies undertaken alongside trials to be identified effectively. The database indexes records by study design, and this field can be easily searched for multicentre or multinational economic evaluations carried out alongside clinical studies.

We therefore searched a subgroup of the NHS EED consisting of economic evaluations carried out alongside single clinical studies (i.e. category A studies on the database). These include RCTs, non-randomised trials with concurrent controls, cohort studies, case–control studies, studies with historical controls, before and after studies in single groups of patients and case series. The search strategies for the NHS EED and the CAIRS software are reported in *Box 1*.

Data extraction instrument

For the purpose of the present review, we developed a data extraction form, the items on which were based on insights from the reviews reported in Chapters 3 and 4. The instrument was divided into three main sections. The first section used 25 items to assess whether a study report provided sufficient information for a decision-maker to be able to establish the relevance of the study to his/her own specific setting. The second section contained 10 items formulated to identify any analytical method that had been adopted to explore the extent to which the economic results varied (a) between the centres/countries participating in the study and (b) between trial centres and routine practice. Given the uncertainty regarding the most appropriate methodology to investigate variability in economic results by location, the data extraction form was designed with an emphasis on describing which techniques were used rather than establishing whether 'good practice' had been adhered to. The final section of the data extraction instrument focused on the

methods used to adapt the results of the trial-based analysis to contexts outside the trial, where these could relate to non-trial locations or non-trial practice.

Data extraction methods

Data extraction was undertaken on the following basis. The included papers were divided into two and a reviewer extracted data on papers in each half (AM and FP). Checks were conducted by each reviewer on 10% of the other reviewer's sample. Data were entered into a database for analysis using STATA 7.0.[165]

Results

Of the initial sample of 107 studies initially identified, 101 satisfied the inclusion criteria and were, therefore, taken into consideration (a full list of the references included in the review is reported in Appendix 3). Among the six rejected studies were four decision analytic models, one 'before and after' study and one cohort study. Of the entire sample, 72% of the economic studies were conducted alongside multicentre (single-country) trials and the remaining 28% were carried out using data from multinational trials. CEA was the most popular form of economic evaluation in the sample (40%), followed by cost–consequence analysis (28%) and cost-minimisation analysis (25%). CUAs (5%) and cost–benefit analyses (6%) were less prevalent in the sample. These numbers sum to more than the total in the sample because some papers conducted more than one form of economic analysis.

Data presentation and reporting

Basis of sampling

On the basis that one of the main advantages of multicentre and multinational trials is that they can produce evidence across a more representative group of patients than single-centre studies, the extent to which this feature was constrained by any form of subsampling on any data collected in the trial was assessed. For this purpose, the review ascertained whether resource use, clinical outcome or quality of life (QoL) data had been collected in only a subset of patients or centres (countries) participating in the study and then extrapolated to the remaining centres (countries).

Among the studies included in this review, 20% were characterised by some form of sampling for resource use. Of these, nine studies used a subgroup of patients, six looked at resource use from a subgroup of centres, one international trial collected data from only one country, three studies used subgroups of patients from different centres and one study gathered data from outside the trial. The collection of clinical outcome data was characterised by sampling to a lesser degree, with only 5% of the studies obtaining data from a subset of patients and 3% of them collecting data from a subset of centres. Among those studies that collected healthrelated QoL, no sampling was undertaken.

Reporting of numbers of centres, patients and countries/jurisdictions

Another element that needs to be reported for a decision-maker to establish the relevance of a study to his/her jurisdiction is the number of centres, countries and patients from which data were collected for the economic evaluation. The larger the overall sample size in the study, the more closely the patients in the trial would be expected to reflect the characteristics of the population from which they were drawn. Furthermore, the larger the number of centres (countries) and the greater the number of patients in each of those locations, the greater is the opportunity to assess the variability in cost-effectiveness results between geographical locations.

In the sample of studies identified, all studies reported the number of patients enrolled in the trial and 98% stated the number of countries participating in the study. Rather fewer (82%) reported the number of centres in which cost and effectiveness data were collected. Very few studies, however, reported the number of patients treated in each centre.

Alternatives described and justified

To establish to what extent the results of a particular study are relevant to a specific jurisdiction, it is essential for the alternative programmes/interventions being compared to be adequately described and justified. In some circumstances, for example, either the intervention or its comparator will not be available in particular locations. In the sample of studies included in this review, only 77% provided a clear description and justification of the alternative interventions being compared, whereas the rest of the sample only provided a description of the comparator used.

Inclusion/exclusion criteria

A clear account of the inclusion and exclusion criteria in a study will help a decision-maker to identify whether the trial patients' clinical and

socio-demographic status is representative of those in their jurisdiction. In addition, this information enables the reader to assess the generalisability of the results to the eligible population, which can be different to the study population. Overall, 77% reported the inclusion and exclusion criteria for the study population and 71% described the baseline characteristics of the patient sample. The most frequently reported were age (65%), gender (56%), disease severity (38%) and co-morbidities (24%). However, only two studies[166,167] assessed the representativeness of the study sample to the study population in the trial centres.

Study perspective

Decision-making in healthcare occurs at different levels. This means that an economic evaluation adopting a societal perspective might not be the preferred viewpoint for a local health authority concerned with maximising the health benefits of its own specific population from its available budget. Similarly, in different countries, budgetary responsibilities may fall on different institutions, and it might be important to be able to disentangle the institutions which incur particular costs and which individuals accrue benefits. In this sample of studies, the economic study perspective was defined in only 42% of the studies and only 36% of these conducted an analysis consistent with the stated perspective.

Study setting

Healthcare delivery varies between different locations. This relates to the specific interventions being directly compared in a study, which is why a full description of those programmes and technologies is important (see above). This type of variation also exists in the more general context of care such as the type of staff undertaking particular interventions and the diagnostic facilities available. Although it can never be exhaustive, a description of the study setting helps decision-makers assess the extent to which a study's results are exchangeable to their own setting. In the sample of studies, some general detail on the study setting was provided by 90% of the sample (e.g. primary care practices, acute hospital treatments). However, specific features of the setting which may be important were less frequently detailed. For example, only three reports[168–170] described the characteristics of the healthcare system(s) where the analysis was conducted. In addition, in just 7% of the sample were one or more centre-specific characteristics reported: professionals' experience,[171] teaching hospital,[172–174] size of the centre[166,171,175] and volume of cases.[171,176]

Unit costs

To estimate the cost of an intervention, trial-based economic evaluation collects patient-level data on resource utilisation. In a multilocation trial, these are typically valued partially or wholly using a single vector of unit costs which may be taken from a single national source or an average over several national sources. Rarely are the resources consumed by a given patient valued using unit costs estimated in the particular centre in which that patient was treated. The generally weak association between the location in which resource use data were collected and the source of the unit costs can complicate the interpretation of the cost results of a multilocation trial **in a particular decision-making jurisdiction**. Hence it is important for the unit costs used in an analysis to be as transparent as possible.

The sample of studies exhibited some encouraging features: almost every study in the sample reported the currency (99%) and the sources of unit costs used to value resource use (92%), although only 75% of the sample provided information regarding the year of costing. Less positively, only 25% of the sample reported resource use and unit costs separately, with the rest of the sample reporting them partially separately (31%) or combined (45%). Furthermore, half of the papers where currency conversion was used (37%) failed to give details of methods and sources for conversion.

Preference data

In CUA, health outcomes are weighted using either patient-specific preferences or utilities elicited from a sample of the population. It is important for the decision-maker to be able to verify the source of preference data. In addition, because preferences are affected by a number of factors (e.g. sociological, cultural) it is important to know the geographical location of the preference sample. In the sample of studies included in the present review, only two of the five CUAs[34,177–180] specified the source of the preference data. However, the other three studies only reported the relevant reference from which the source of the valuation data could be ascertained. For four out of the five studies, public preferences were used using the EQ-5D (n = 2) or the Rosser valuation system; one study used several instruments (including EQ-5D). The final paper used patients' preferences directly elicited using a rating scale.

Analytical methods relating to generalisability within the trial

The second part of the data extraction process focused on the analytical methods which studies

have employed to assess the generalisability of their results across study centres, in other words, the extent to which variability between centres was formally assessed. It is important to distinguish the use of more formal quantitative methods to investigate variability in results by location (statistical techniques, such as regression analysis) from the use of sensitivity analysis to test the implications of changing the value of particular parameters (e.g. unit costs) for study results. The former can be used explicitly to assess the extent to which variability in costs, effects or cost-effectiveness can be explained by patient-level covariates (e.g. case mix) or the location in which the patient was treated. Sensitivity analysis, by contrast, is often used as a way of assessing the importance of **parameter uncertainty** (i.e. **precision**) for the results of a study. However, to the extent that some parameters are known systematically to vary by location (e.g. unit costs), this form of analysis can also be helpful to assess the implications of parameter **variation**. In addition, some studies provide a more qualitative report of the factors that may influence the generalisability of the results without producing any formal quantitative assessment.

Statistical methods used to explore variability in the results

Very few studies attempted to explore the variability of their results by location using statistical methods and only two studies used regression methods to assess variability. In the first of these,[181] the authors observed a difference in the treatment effect across different countries and explored the possible determinants of this variability. The analysis showed that the geographical variability was not related to the intervention itself. In the second study,[133] the authors conducted an ordinary least-squares regression of the pooled logarithmically transformed costs to explore whether these were varying from country to country. The regression included covariates of trial arm, country dummy, age, gender, smoker, employment status and two clinical predictors of disease severity. The study found no statistically significant difference in costs between the two treatment groups, but it did find statistically significant differences in costs between the countries. The authors did not explore interactions or relate costs to outcomes in their regression analysis.

Multinational trials

There is interest in looking specifically at the analytical basis of the economic evaluations alongside multinational trials. With these studies, the ways in which potential variability in costs, effects and cost-effectiveness is handled are particularly important given that between-country variation is likely to be more pronounced than within-country variation.

In the sample of 28 multinational studies (see *Table 11*), nearly 50% identified a single vector of unit costs, from one country, and applied these costs to pooled resource use from all countries in the trial.[34,86,179,182–191] The next most popular method (four studies) was to calculate several different unit cost vectors for different countries, and to apply each of these to all resource use data in the trial.[78,173,192,193] In effect, this amounts to a series of sensitivity analyses based on different countries' unit costs.

Only one study valued patients' resource use using the unit costs estimated in the country where that patient was randomised.[133] One other study used a

TABLE 11 *Details of the basis of the cost analysis in economic evaluations undertaken alongside multinational trials*

Type of analysis	Number of trials (%)
1. Single set of unit costs applied to resource use in all locations	13 (46%)
2. Several vectors of country-specific unit costs (taken from different countries) applied to all resource use in separate analyses	4 (14%)
3. Estimates of resource use and cost taken from outside the trial and from one country related to effect data from the trial	2 (7%)
4. Resource use and unit costs taken from one country in the trial and related only to that country's effects	2 (7%)
5. Combination of 1 and 4	2 (7%)
6. Country-specific unit costs applied to country-specific resource use.	1 (4%)
7. Average vector of unit costs from several countries applied to resource use in all locations	1 (4%)
8. Combination of country-specific unit costs applied to resource use in that country for 6/11 countries and average unit costs applied to all resource use in the other 5 countries	1 (4%)
9. Combination of 4 and 6	1 (4%)
10. Unclear	1 (4%)

TABLE 12 *Sensitivity analysis approach by type of variables that can be affected by location*

Variables potentially influenced by location	Qualitative sensitivity analysis (n = 5)	Sensitivity analysis based on ad hoc ranges (n = 40)	Sensitivity analysis based on empirical ranges (n = 8)
Compliance	0	2	0
Case mix	0	0	1
Epidemiology	0	0	0
Opportunity costs	0	1	0
Utilities	0	1	0
Demographic factors	0	0	1
Clinical practice	0	8	4
Organisation	0	1	0
Learning effect	0	1	1
Relative prices level	1	9	2
Other variables	4	17	0

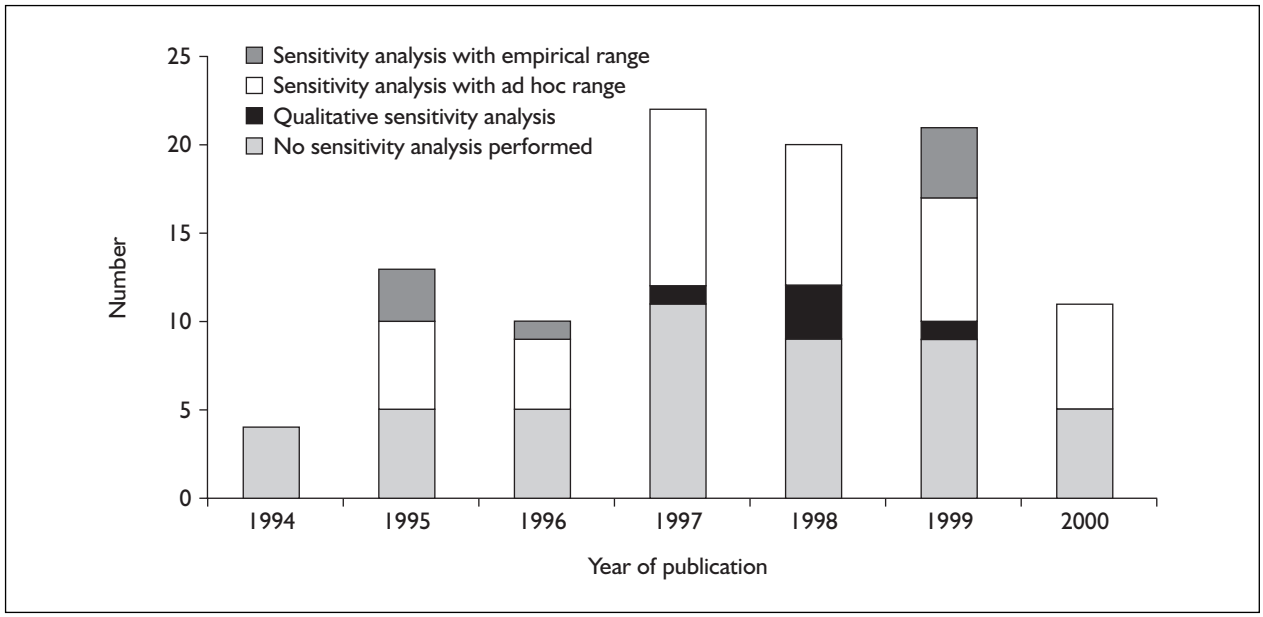

FIGURE 1 *Type of sensitivity analysis by publication year*

mixture of country-specific unit costs to value resource use in six out of 11 countries in the trial and the average across those six countries to value resource use in the other five countries.[181] A range of methods was used in the other studies in the sample. All but two studies related their cost estimates to effectiveness data pooled across all countries. These two exceptions isolated one country from the trial and estimated costs, effects and cost-effectiveness solely using patient data from that country.[194,195]

Sensitivity analysis

Table 12 describes the different sensitivity analysis approaches used in the studies included in the present review. Details of those variables, which have the potential to vary by location, which were assessed in these analysis are also provided. This table looks at the use of sensitivity analysis identifying three main approaches: (a) qualitative (i.e. no formal analysis); (b) sensitivity analysis informed by *ad hoc* ranges of variation for the key parameters (i.e. max.–min., guess-estimates, ±50%); and (c) sensitivity analysis using ranges observed in the data (e.g. 95% confidence intervals). Finally, *Figure 1* illustrates the distribution of type of sensitivity analysis by year of publication.

Almost half of the studies in the sample (n = 48) did not provide any assessment which could help a reader assess the generalisability of their results by location. Two multinational studies (discussed above) reported sensitivity analyses which were

explicitly focused on variability by location.[133,181] Of those studies undertaking sensitivity analyses to look at the robustness of their results more generally (i.e. not specifically to assess variation by location), there were five examples of qualitative sensitivity analysis, of which only one identified a variable with potential to vary by location. A total of 40 studies explored the sensitivity of their results using *ad hoc* ranges for one or more of the parameters of interest. In this subsample, the sources identified to have an impact on their results, and which may systematically vary by location, were absolute/relative costs or prices (n = 9), clinical practice (n = 8) and compliance (n = 2). Very few studies (n = 8) based their sensitivity analyses on a range empirically established from the data (e.g. 95% confidence interval) and, as emerges from *Figure 1*, these tend to be more recently conducted analysis.

Of the original sample, only 26% compared their results with those of similar studies, and five of these set their results in the context to other independent interventions (e.g. in a league table). A number of studies not comparing their results with any other previously published result claimed that their report was the first in that clinical area.

Analytical methods to adjust trial results to non-trial context(s)

Finally, there were no examples of the application of methods to adjust trial results from those based on patient-level data collected within the trial to non-trial contexts in terms of location, study to practice or over time.

Discussion

Study reporting

An important focus of this review has been on how economic evaluations alongside multilocation trials report their results. This is not to judge whether the studies in the review are generalisable in themselves, but rather to establish whether enough information is provided to assess their generalisability. This is because many decision-makers will be focused on the results of a study as they apply to their particular jurisdiction which may be all centres in trial undertaken in a single country, one centre (or a proportion of centres) in a multilocation trial or one country in a multinational trial. Therefore, some decision-makers may be interested in all the locations in which the data in the trial were collected. In this situation, it is important to be able to establish how representative trial patients are of the more general population from which they are drawn and for which the decision-maker is responsible. Other decision-makers may be directly interested only in particular locations in the trial – for example, a UK decision-maker is focused on the implications of a multi-national trial for UK hospitals. Hence an important feature of any study is the extent to which it provides sufficient information for a decision-maker to assess the relevance of its results to a particular location or set of locations.

The way in which economic studies are reported has been a key feature of general guidelines for good quality in economic evaluation which have been available to authors for some years.[162] The availability of this guidance is partially reflected in some encouraging results here regarding the reporting of those study characteristics which will influence perceived relevance in particular locations. For example, although there remain a worrying proportion of studies which are not following widely accepted reporting standards, at least the majority of studies describe and justify the alternatives being compared (77%) and the inclusion and exclusion criteria employed (77%).

However, other findings in this review indicate that some of the basic features of good reporting are not being widely adhered to. The study perspective was defined in only 42% of the studies in our sample. Given the importance of perspective as a basic method in economic evaluation, this is a surprising result. If analysts neglect to provide this sort of information, the onus is on the decision-maker to interpret the perspective from available information on costs and effects. A surprising result was that only 25% of the studies reported resource use and unit costs separately, with the rest of the sample reporting them partially separate (31%) or combined (45%). The separate reporting of unit costs and resource use is widely considered 'good practice'[1,3] and a failure to do this prevents the decision-maker from assessing the extent to which the vector of unit costs used in a particular study are relevant to his/her jurisdiction.

Some other aspects of the observed reporting provide further barriers to the assessment of a study's generalisability and relevance to particular jurisdictions, but these have not been so widely reflected in general reporting guidelines for economic evaluation. First, only two studies[194,195] assessed the representativeness of the study sample to the study population in the trial centres. If only a small proportion of patients are recruited

into trials from amongst those eligible (i.e. fulfilling the inclusion and exclusion criteria), and the characteristics of those included are not representative of the population, the generalisability of the study would probably be compromised. Second, in 91 articles, a definition of the study setting was provided, but only three papers described the characteristics of the healthcare system(s) where the economic analysis was conducted.[168–170] In addition, only seven studies reported one or more centre-specific characteristics such as the volume of cases they treat.[166,173–176,196,197] These results are perhaps as expected, and it can be argued that authors would not normally be expected to provide such information. However, any reasonable assessment of the extent to which results found in a combination of locations apply to a particular location (within or outside the trial) requires information on what types of centres were included in a study and how trial patients compare with non-trial patients.

Further development of what constitutes 'good reporting' is required. In the past, the space constraints in journals (particularly clinical ones) have placed limitations on the sorts of detail that a decision-maker might need to assess generalisability. However, some journals now provide an opportunity to place more information on a study on their website. Technical reports can also be made available (via the authors) to extend the amount of information about the study.

Analytical methods

In recent years, statistical approaches have emerged in the methods literature to assess systematically whether costs and/or cost-effectiveness vary by location[25,33,34,131,133–135,138] (see Chapter 4). However, in the sample of studies reviewed here, the use of these types of methods was sparse. There may be a number of reasons for this. First, a large part of the studies included in this review was published between 1995 and 1998 when some of the methodological contributions in this field were still to be published. Second, it is likely that editorial needs may have limited the scope for exploratory work within the main economic paper. For example, the main economic analysis of a trial of trilazad mesylate for aneurysmal subarachnoid haemorrhage has been reported in a non-economics journal[181] and further methods work in economics journals.[25,198] Finally, the focus of the paper may be clinical and the economic analysis was a secondary consideration.

The review found no studies which had used analytical methods to extend the within-trial results to other locations. This may be because each original trial was designed to address a particular question relating to one or more locations and there is little interest in extrapolating outside trial locations. Alternatively, it could be because there is an implicit assumption that results do not vary between locations (centres or countries) or that trial locations are representative of non-trial locations. The first of these may be true of some types of clinical measurements but, as reviewed in Chapter 3, is unlikely to be accurate for economic data. An assumption that trial locations are representative may also be doubted, given that not all centres treating patients are research focused. The process of adapting results to non-trial locations is probably more likely to be undertaken using a decision modelling framework as described in Chapter 4.

The results of this review seem to be consistent with an earlier review conducted to consider variation in study results between locations. Walker and Fox-Rushby conducted a review of economic evaluations of control strategies against parasitic diseases in terms of internal and external validity.[199] Only seven studies explicitly discussed the generalisability of results to other settings. Four of these provided blanket statements advocating the implementation of a new intervention without any discussion as to the applicability of their results to different economic or epidemiological settings. In three studies, generalisability was examined through the use of itemised cost menus in which the unit costs of inputs could be changed to reflect those of other countries or regions.

Conclusions

Given the expectation that costs and cost-effectiveness can vary by geographical location, there is a need to appreciate this source of variability in analysing and reporting economic evaluations undertaken alongside multicentre and multinational trials. This review indicates, however, that few studies provide sufficient evidence for decision-makers to establish the relevance or to adjust the results of the study to their location of interest. It is also clear that few studies published during 1992–2000 were utilising statistical methods formally to assess the variability in results between locations.

Chapter 6

Case study I: assessing generalisability in trial-based economic evaluation using multilevel modelling

Chapter 3 showed that, for a range of reasons, there is likely to be variability in cost and cost-effectiveness results between locations. Chapters 4 and 5 showed that, in the context of trial-based economic evaluation, the conventional analytical approach to this situation has been to assume that variability between locations relates only to prices/unit costs and that resource use and effectiveness measures are exchangeable between locations. For example, Sculpher and colleagues applied UK unit costs to resource use data collected in a multicentre/multinational trial and related costs to clinical effects measured in all countries.[155] Some studies have made an attempt to adjust resource use and effects, in addition to unit costs, to relate to specific locations. However, as shown in Chapter 2, this has typically been undertaken using decision models where non-trial data are used to extend trial results to non-trial locations.[31] As indicated in Chapter 5, trial-based economic studies have rarely sought to address variability between location simultaneously in all forms of data (resource use, unit costs and effects) using data collected only in the trial.

Chapter 4 reviewed two methodological studies which have attempted to develop methods to look systematically at variability between locations in cost-effectiveness studies alongside trials. In the context of a multinational trial in subarachnoid haemorrhage, Willke and colleagues used regression analysis to model separately the effects of treatment on costs and outcomes and of outcomes on cost.[25] The use of a treatment–country interaction term allowed for the estimation of country-specific cost-effectiveness ratios based on country-specific resource use and outcomes as measured in the trial, and each country's unit costs. The considerable extent of variation between countries in cost-effectiveness indicated the potential lack of generalisability of the trial's overall (i.e. pooled) results.

Another approach, suggested by Cook and colleagues,[131] is based on a test for heterogeneity that should inform the decision regarding whether data from different centres/countries can be pooled into a single analysis or if, instead, separate analyses by location are necessary. This method is analogous to clinical tests of heterogeneity and relies on the identification of 'statistically significant heterogeneity' to define whether the overall results of studies are generalisable.

This chapter seeks to contribute to analytical methods in this area by describing and illustrating the use of multilevel modelling (MLM) as a means of assessing the variability and generalisability of cost-effectiveness across locations. The premise of these methods is that the variability in costs and outcomes which would be expected between locations is (at least in part) a manifestation of the hierarchical structure of the data collected in trial-based economic evaluation. In these trials, patients are naturally clustered within centre and, in multinational trials, it is not unusual to observe patients clustered within centres within countries. This 'clustering effect' has the potential to exist whether the patient or the centre represents the unit of randomisation. If data are significantly 'clustered' in this way, then failing to use statistical methods which reflect this characteristic can generate misleading results.

Recent work on regression methods in economic evaluation opens up a range of methodological options because a single-dimensional measure of cost-effectiveness (net benefit) can be included as a dependent variable within regression models.[156] The use of regression methods in this context represents a major potential improvement in analytical methods, as it is possible to estimate the cost-effectiveness of an intervention while simultaneously controlling for differences in baseline covariates, in addition to facilitating the use of subgroup cost-effectiveness analysis. As a further development within this general regression framework, MLM represents a means of explicitly modelling the hierarchical nature of cost-effectiveness data. MLM has been used elsewhere in economic and social research,[200,201] but no examples of its application to economic evaluation in healthcare have been identified.

This chapter is structured as follows. The next section focuses on the principles of stochastic cost-effectiveness analysis and the use of net benefit regression. The role of clustering in economic evaluation alongside clinical trials, where randomisation occurs at patient level, is discussed in the following section. The trials which are used to illustrate MLM are then introduced; these methods are employed to analyse the trial data and the results are compared with those obtained using standard approaches to cost-effectiveness alongside trials. The chapter concludes with a discussion of the potential uses of this new approach attempting to draw some future lines of research in this field.

Stochastic methods and net benefit regression

From deterministic to stochastic cost-effectiveness analysis

Standard methods of trial-based economic analysis estimate total costs per patient, simply multiplying the volume of resources used by each patient by the unit costs for those resources and summing these over the entire study period. Similarly, health state valuation data – based on direct valuation or on responses to preference-based instruments such as the EQ-5D[202] – are collected at the individual level. These data are used to calculate patient-level QALY estimates over the entire study period. In other words, these mean cost and mean outcomes results are estimated with no allowance made for the potential clustering of costs and outcomes by location and no adjustment undertaken for baseline covariates.

It is only in recent years that trial-based economic analysis has moved away from deterministic methods, where only point estimates of mean cost and benefits of the alternative interventions are reported, to stochastic approaches to data analysis. With the latter, patient-level data are used to quantify the sampling variation associated with costs and outcomes and formally to reflect this in measures of uncertainty around estimates of mean costs, effects and cost-effectiveness. Key contributions in stochastic methods include the most appropriate methods to estimate confidence intervals (CIs) around mean incremental cost-effectiveness ratios (ICERs)[203–205] and how to present and characterise sampling uncertainty in a way that is useful to inform the decision-making process.[206,207]

Net benefits

This methodological work indicated that there are two important problems associated with representing the stochastic nature of the ICER. The first is the interpretation of negative ICERs, which can mean that the new intervention, for example, is either more costly and less effective (dominated) or less costly and more effective (dominant) than one or more comparators – two very different situations. This means that the calculation of CIs for the ICER when the joint density of differential cost and effects lies across several quadrants in the cost-effectiveness plane is not straightforward.[208] The second problem relates to the quantification of the sampling uncertainty around the ratio statistic when there is a non-negligible probability that the denominator takes values close to zero.

Moving the stochastic analysis of patient-level data into the net benefit framework has been suggested as a mean of avoiding these problems with ICERs.[151,209] This is implemented by reinterpreting the traditional ratio-based decision rule:

$$ICER = \frac{\overline{C_1} - \overline{C_0}}{\overline{E_1} - \overline{E_0}} < \lambda \qquad (1)$$

as a linear expression which can be formulated in terms of incremental net monetary benefits [equation (2a)] or incremental net health benefits [equation (2b)]:

$$INMB = (\overline{E_1} - \overline{E_0})\lambda - (\overline{C_1} - \overline{C_0}) > 0 \qquad (2a)$$

$$INHB = (\overline{E_1} - \overline{E_0}) - \frac{\overline{C_1} - \overline{C_0}}{\lambda} > 0 \qquad (2b)$$

where $\overline{C_1}$ and $\overline{E_1}$ are, respectively, the mean cost and effect of the new intervention and $\overline{C_0}$ and $\overline{E_0}$ are the mean cost and effect of the standard treatment. The parameter λ represents the decision-maker's maximum willingness to pay to achieve an additional unit of effectiveness and can be interpreted as the shadow price of the budget constraint (that is, the opportunity cost of the new intervention in terms of what the health system would need to give up in order to fund it) or some measure of value that a population would attach to a unit of health outcome.

Once λ has been defined, the adoption of the net benefit approach offers an unequivocal decision rule as it allows the value for money of the alternative strategies to be clearly determined. Furthermore, quantifying the sampling uncertainty around the mean incremental net

benefit becomes straightforward as the linearity of the net benefit statistic helps to overcome the problems suffered by the ICER.[209]

When individual patient data on costs and effects exist, as in trial-based studies, net monetary benefits (NMB_i) and net health benefits (NHB_i) can also be defined at the level of the individual patient (i) as shown in equations (2c) and (2d). These are obviously absolute rather than incremental and relate to the treatment to which the patient was allocated.

$$NMB_i = E_i\lambda - C_i \quad (2c)$$

$$NHB_i = E_i - \frac{C_i}{\lambda} \quad (2d)$$

Cost-effectiveness acceptability curves

The cost-effectiveness acceptability curve (CEAC) has become widely used in trial-based economic evaluation as a means of representing the stochastic uncertainty in cost-effectiveness explicitly, and emphasising the importance of λ in the decision rule about a preferred intervention.[206] The CEAC has a naturally Bayesian interpretation,[210] as it shows the probability that expected (mean) net benefits are greater for one treatment than another given the data in the trial, and this is usually how they are presented in applied studies.[155,211] However, it can also be adopted as a way of presenting cost-effectiveness data in a Frequentist framework.[210]

Net benefit regression

As a natural extension of the stochastic framework in the analysis of patient-level data, Hoch and colleagues have proposed a reformulation of the cost-effectiveness problem within a standard regression framework.[156] Regression techniques can improve the range of research questions that can be investigated in stochastic cost-effectiveness studies and can help to explore further the different dimensions of uncertainty in the results. In particular, expected net benefits associated with a given treatment can be estimated more precisely by controlling for other covariates such as baseline socio-demographic and clinical characteristics of patients. Indeed, in observational studies this process of controlling for potentially confounding variables is an essential step in analysing the data. By exploring the effect of interactions between a particular treatment and baseline covariates, the cost-effectiveness of interventions in particular subgroups of patients can also be estimated.

The rationale for the adoption of a regression approach in healthcare economic evaluation stems from the linearity of the net benefit statistic. Using a model regressing patient-level net monetary benefits (NMB_i) against the treatment arm dummy variable (t_i), Hoch and colleagues demonstrated the equivalence of the regression-based approach with a 'standard' CEA. This is illustrated as

$$NMB_i = \beta_0 + \beta_1 t_i + \varepsilon_i \quad (3)$$

In this formulation, the net monetary benefit for the ith patient in the trial is the patient-level net benefit defined in equation (2c). The coefficients β_0 and β_1 are, respectively, the intercept and the slope term obtained from a standard ordinary least-squares (OLS) regression. The term ε_i is a zero-idiosyncratic error term with constant variance. It is often assumed that ε_i is normally distributed.

In terms of the interpretation of the results from the OLS regression, the estimated coefficient β_0 represents the mean net benefit in the group receiving the standard treatment in the trial, the sum of the two estimated coefficients ($\beta_0 + \beta_1$) is the mean net benefit in the new treatment arm and β_1 is the incremental net benefit between the two arms of the trial. In the model presented here, these estimates are identical with that which would be obtained from a standard CEA. As noted above, the model described in equation (3) can be extended to include patient-level covariates to control for any factor thought to confound the estimate of the treatment effect and treatment–covariate interactions to assess cost-effectiveness in patient subgroups.

Dealing with clustered data

The problem

As has been pointed out in earlier chapters, it could be argued that, in the context of multicentre or multinational RCTs, individuals within a specific location are more similar to each other than individuals from different locations, in particular in terms of the type of clinical management they receive. For example, there may be differences between hospitals in how they manage patients, and this may manifest itself in trials.[212] In addition, there may be differences between centres in the case mix of patients which are referred to them. In other words, patients are inevitably clustered within centres and, consequently, the data will be characterised by an inherent hierarchical structure.

In trials which randomise by centre rather than by patient (cluster-randomised trials), the hierarchical nature of the data is an inevitable implication of the design of the study. However, at least for economic analysis, the characteristic is also likely to exist for trials where the patient is the unit of randomisation. In these cases, one of the assumptions underpinning the use of standard OLS regression – that the random errors, ε_i, are independently distributed – does not hold, as observations within clusters will be correlated. Failure to acknowledge this feature of the data in the analysis will result in exaggerated precision in estimates of the incremental net benefit associated with a treatment. More formally, without adjustment for clustering, standard OLS can produce inefficient parameter estimates and incorrect standard errors.

The intraclass correlation coefficient

The extent to which OLS will produce incorrect inference depends on the extent of clustering in the data. The degree of clustering in a hierarchical dataset is generally measured by the **intraclass correlation coefficient (*ICC*)**, a statistic summarising the degree of dependency in nested observations. In a simple two-level nested data structure (for example, patients clustered within hospitals), the concept of *ICC* can be illustrated with the help of a random effects analysis of variance (ANOVA) model[213] described by the equation

$$NMB_{ij} = \beta_0 + u_{0j} + \varepsilon_{ij} \quad (4)$$

where β_0 is the average *NMB* from pooling all the observations in the data set, u_{0j} is the specific effect of the jth 'centre' (e.g. the hospital) and ε_{ij} is the residual effect for the ith patient within the jth centre. Unlike the expression in equation (3), *NMB* are now expressed with two subscripts, i and j, which illustrate the individuals' (i) membership of specific groups (j). The mean *NMB* for a particular centre, k, can then be expressed as $(\beta_0 + u_{0k})$, with each individual-patient observation departing from this group mean by a random value ε_{ij}. Notice that this model has two random components, u_{0j} and ε_{ij}. Both are assumed to have zero mean and are uncorrelated; u_{0j} has variance ${\sigma_{u0}}^2$ and ε_{ij} has variance ${\sigma_\varepsilon}^2$.

In this model, the quantity β_0 is often referred to as the fixed part of the model, whereas u_{0j} and ε_{ij} constitute the random parts of the model. The quantities of interest, which are estimated from the data, are β_0, ${\sigma_{u0}}^2$ and ${\sigma_\varepsilon}^2$. The last two quantities are often referred to as variance components and are used to derive the *ICC* as follows:

$$ICC = \frac{{\sigma_{u0}}^2}{{\sigma_{u0}}^2 - {\sigma_\varepsilon}^2} \quad (5)$$

where the *ICC* can be interpreted as the proportion of the total variance that can be attributed to centre-level variation. The *ICC* can take values of between 0 and 1 inclusive. Quantification of the expected *ICC* prior to conducting the study is essential in the design phase when determining the appropriate sample size, even in trials where the randomisation occurs at patient level.[214,215] In addition, the *ICC* is fundamental in the analysis phase when making inference and prediction from data.

Multilevel modelling in economic evaluation using patient-level data

Overview

In situations where some form of natural clustering in the data exists and where the *ICC* is greater than zero, the use of OLS regression may be inappropriate and can lead to spurious conclusions. In particular, the greater the *ICC*, the less appropriate is the use of OLS analysis (the higher the *ICC*, the stronger is the correlation between the idiosyncratic error terms and the greater the violation of OLS assumptions). In these situations, MLM is the appropriate form of analysis. This is a statistical regression-based technique which takes into account the hierarchical or clustered nature of the data. This technique has been used in different fields such as sociology,[201] education,[200,216] epidemiology[217] and economics.[218] Applications of multilevel modelling include the analysis of longitudinal data, where the interest remains on the dynamics of repeated observations nested within individuals over time,[219] the analysis of cross-sectional surveys in which the researcher has multiple observations nested within time periods,[220] meta-analysis[221] and multivariate response.[222] The uses of MLM in health economics has also been reviewed.[21] Although the authors discussed the potential of the method in economic evaluation, no application has been identified in the literature.

Variance component (random intercept) specification

To describe MLM in more detail, level-1 units are used to describe the micro or individual level units in the study and level-2 units refer to the macro

units or groups. The definition of macro and micro units is, of course, contextual in any analysis. If the focus, for instance, is to study patients within hospitals, then the former will be micro units whereas the latter are macro units. However, if we were to study hospitals' performances within different regions in a country, then hospitals would be our micro units and regions our macro units.

MLM allows the simultaneous consideration of the **within**-group (e.g. within hospitals) and **between**-group (e.g. across hospitals) variability of the dependent variable. The partitioning of unexplained variation into these two components can be done once the model has been conditioned on a set of explanatory variables. In general terms, we define N as the number of groups in the study (e.g. centres) and n_j as the number of individuals within the jth group ($j = 1, 2, 3, \ldots, N$). The index i represents level-1 observations within the jth group ($i = 1, 2, 3, \ldots, n_j$). The total sample size is

$$M = \sum_j n_j$$

If we observe NMB for the ith patient within the jth group, then we can re-express the regression model presented in equation (3) as a simple multilevel random intercept model as follows:

$$NMB_{ij} = \beta_{0j} + \beta_1 t_{ij} + \varepsilon_{ij} \quad (6)$$

where β_{0j} and β_1 are the intercept and the slope, respectively. In a simple variance component specification, β_{0j} is characterised as follows:

$$\beta_{0j} = \beta_0 + u_{0j} \quad (7)$$

where β_0 and β_1 are the population average intercept and the average slope, respectively, with u_{0j} representing the jth group departure from the average intercept. Notice that the coefficient β_1 has only one subscript, which means that the slope is assumed to be fixed across groups. Substituting equation (7) into equation (6) yields,

$$NMB_{ij} = \beta_0 + \beta_1 t_{ij} + u_{0j} + \varepsilon_{ij} \quad (8)$$

The model comprises a fixed and a random part. The difference with respect to the standard OLS specification lies in the fact that the latter includes only one random component [i.e. the error term ε_i in equation (3)], whereas the multilevel model can have two or more random elements. In Equation (8), the **fixed** part of the model is given by the term $(\beta_0 + \beta_1 t_{ij})$, whereas $(u_{0j} + \varepsilon_{ij})$ is its **random** part. The term u_{0j} represents the variability across level-2 units and is assumed to be independent and identically distributed with zero mean and *a priori* unknown variance. In addition, u_{0j} is assumed to be independent of the level-1 residuals, ε_{ij}, which are assumed independent and identically distributed. It is usual to assume that

$$u_{0j} \sim N(0, \sigma_{u0}^2) \quad (9)$$
$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

The conditional variance of the NMB_{ij} is therefore given by

$$Var(NMB_{ij}) = \sigma_{u0}^2 + \sigma_\varepsilon^2 \quad (10)$$

With respect to the interpretation of the results, in a hypothetical RCT of patients clustered within hospitals, the regression terms can be interpreted as follows. The term β_0 [in equation (8)] represents the average net benefit for the group of patients receiving the standard treatment, with u_{0j} reflecting the departure from the average net benefit for the jth hospital. Similarly, the regression coefficient $(\beta_0 + \beta_1)$ corresponds to the average net benefit for the group receiving the new intervention. The coefficient β_1 is the average incremental NMB. Notice that, in the model specification presented in equations (8) and (9), the average net benefit for the standard treatment and the new intervention arm are allowed to vary by centre, but their difference (i.e. β_1) does not vary by centre. In other words, in this model the implicit assumption is that the average incremental net monetary benefit ($INMB$) does not vary by location.

Finally, the ICC in the simple variance component specification is

$$ICC = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 - \sigma_\varepsilon^2} \quad (11)$$

Random effects specification

For cost-effectiveness analysis using patient-level data from a number of centres and/or countries, a major interest is in the variability of average $INMB$ (i.e. the incremental cost-effectiveness of a treatment relative to a comparator) across different locations.[25] Therefore, it is necessary to move from a simple variance component specification, such as that in equation (8), to a full random effects model (i.e. random slope and random intercept). This is done by specifying the intercept and slope coefficients as

$$\beta_{0j} = \beta_0 + u_{0j} \quad (12)$$
$$\beta_{1j} = \beta_1 + u_{1j}$$

where

$$u_{0j} \sim N(0, \sigma_{u0}^2) \qquad (13)$$
$$u_{1j} \sim N(0, \sigma_{u1}^2)$$

The key difference from the model illustrated in the previous section is that, in this specification, the coefficient β_{1j} had an index j to signify that this coefficient is now allowed to vary across centres, with u_{1j} representing the jth group departure from the population average slope, β_1. The net benefit regression can now be written as

$$NMB_{ij} = \beta_0 + \beta_1 t_{ij} + u_{0j} + u_{1j} t_{ij} + \varepsilon_{ij} \qquad (14)$$

or equivalently

$$\begin{aligned} NMB_{ij} &= \beta_{0ij} x_0 + \beta_{1j} t_{ij} \\ \beta_{0ij} &= \beta_0 + u_{0j} + \varepsilon_{ij} \\ \beta_{1j} &= \beta_1 + u_{1j} \end{aligned}$$

with $x_0 = 1$ for all i and j.

Again, a fixed and a random part compose the model, with the **fixed part** of the model represented by the term $(\beta_0 + \beta_1 \cdot t_{ij})$ and the **random part** being given by the term $(u_{0j} + u_{1j} \cdot t_{ij} + \varepsilon_{ij})$. Notice that the random part of the model now includes the term $u_{1j} t_{ij}$. The components u_{0j} and u_{1j} are the unexplained level-2 effects. These are assumed to be correlated within the same group. The implications of this assumption will be considered later in this section.

The terms u_{0j} and u_{1j} are assumed to be independent between groups and to be identically distributed with zero mean (given the level of the explanatory variable t_{ij}) and *a priori* unknown but constant variance. In addition, these level-2 random effects are independent of the level-1 residuals ε_{ij}, which are assumed independent and identically distributed, $\varepsilon_{ij} \sim \mathrm{N}(0, \sigma_\varepsilon^2)$. More formally, we write

$$\begin{aligned} &Var(u_{0j}) = \sigma_{u0}^2,\ Var(u_{1j}) = \sigma_{u1}^2, \\ &Cov(u_{0j}, u_{1j}) = \sigma_{01},\ Var(\varepsilon_{ij}) = \sigma_\varepsilon^2 \end{aligned} \qquad (15)$$

In this specification, the total unexplained variance of the NMB_{ij} conditional on t_{ij} is

$$Var(NMB_{ij} | t_{ij}) = \sigma_{u0}^2 + 2\sigma_{u01} t_{ij} + \sigma_{u1}^2 t_{ij}^2 + \sigma_\varepsilon^2 \qquad (16)$$

Since $t_{ij} = 0, 1$ for all i and j, then

$$Var(NMB_{ij} | t_{ij} = 0) = \sigma_{u0}^2 + \sigma_\varepsilon^2$$

$$Var(NMB_{ij} | t_{ij} = 1) = \sigma_{u0}^2 + 2\sigma_{u01} + \sigma_{u1}^2 + \sigma_\varepsilon^2 \qquad (17)$$

and, therefore, we can define an *ICC* for each arm of the trial.

The regression terms β_0 and β_1 have the same interpretation as in the variance component specification, with the addition that, in equation (14), not only are the average *NMB* in both arms of the trial allowed to vary by location, but so is the **incremental** net benefit of one treatment relative to the other.

It is worth pointing out that in this case we will have two *ICC*s:

$$ICC = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 - \sigma_\varepsilon^2} \qquad (18a)$$

for the standard treatment arm (i.e. for $t_{ij} = 0$) and the other:

$$ICC = \frac{\sigma_{u0}^2 + 2\sigma_{01} + \sigma_{u1}^2}{\sigma_{u0}^2 + 2\sigma_{01} + \sigma_{u1}^2 + \sigma_\varepsilon^2} \qquad (18b)$$

for the new intervention arm (i.e. for $t_{ij} = 1$).

The model assumption that u_{0j} and u_{1j} are not independent makes intuitive sense in the context of multicentre/multicountry studies. This is because we can expect that, within the same centre, the net benefits of the two a lternative treatments being compared will be correlated. On the cost side, this may be due to the fact that both interventions will be affected by the hospital production function and on the outcome side we can assume that the general level of care in the two arms does not differ apart from the interventions being investigated.

Advantages of multilevel models

There are three main advantages of using the MLM approach that need to be distinguished: parameter estimates, the generalisability of results and location-specific cost-effectiveness.

Parameter estimates

Compared with the OLS method, if the *NMB* of individuals within centres are correlated (i.e. the assumption of independence of observation is violated), multilevel analysis allows more efficient estimates of the *INMB* and the correct standard error to be obtained. This is particularly important when CEACs are being presented to

decision-makers to indicate the level of decision uncertainty associated with a particular allocation of resources. It is also important in quantifying the cost of uncertainty and the expected value of perfect information.[151,153]

The generalisability of results

The estimation process of the multilevel family of models is based on the assumption that the group-level units (here centres/countries) are randomly selected from a population of such units. For this reason, the use of MLM is particularly useful to assess the generalisability by location of the results of economic evaluation studies alongside clinical trials. This can be achieved in two ways. First, MLM can be used to estimate cost-effectiveness results across all locations which allow for the hierarchical nature of the data by adding additional uncertainty to the results. Second, MLM can be used to generate cost-effectiveness estimates which are specific to each location in the study; the extent to which the results of the analysis are consistent across locations can then be established. How MLM is used in a particular study will partly depend on whether the study is a multicentre or multinational trial.

Location-specific cost-effectiveness

Location-specific estimates of cost-effectiveness are facilitated through a feature of MLM. The random effects (or residuals) u_{0j} and u_{1j} in equation (14) are **latent variables** and not statistical, so they are not obtained as part of the parameter estimation process (the parameters of interest are the fixed part coefficients and the variances of the random components). Nevertheless, it can be useful to quantify the residuals and this can be achieved through what is termed **shrinkage estimation**.

Consider the raw residual from a variance components model [equation (8)] and calculated as $r_{ij} = NMB_{ij} - \hat{NMB}_{ij}$ where patients are clustered in centres. The raw residual for the jth centre, r_j, could then be calculated as the mean of r_{ij} over all patients within the centre. Shrinkage estimation of centre specific residuals for the intercept term, for instance, are obtained as follows:

$$\hat{u}_{0j} = \frac{\sigma_{u0}^{2}}{\left[\sigma_{u0}^{2} + \left(\frac{\sigma_{\varepsilon}^{2}}{n_j}\right)\right]} \times r_j \qquad (19)$$

The multiplier in equation (19) is always ≤ 1 so that the estimated residual cannot be greater than the raw residual for a specific centre. This multiplier represents the shrinkage factor. Shrinkage is large when either σ_{ε}^2 is large relative to σ_{u0}^2, or when n_j is small or both. In these circumstances, shrinkage reflects a lack of information about centres and so the raw centre-specific residual is shrunk towards zero. This use of shrunken residuals is useful when calculating centre-specific *INMB*, as the less information, or more uncertainty, relating to an individual centre's results, the more it will be appropriate to rely on the overall mean as the best estimate for that centre.

Case studies

Purpose

In this section, the methods described above are applied to three case studies. First, the results of the analyses conducted using standard economic evaluation methods are presented and discussed. Next, the equivalence of the OLS regression approach to the standard analytical methods is confirmed. Finally, the results obtained using multilevel regression are compared with those obtained with standard analysis and OLS regression. In particular, there will be a demonstration of the use of MLM to obtain (a) more appropriate estimates of the incremental net benefit and of its standard error for the population of groups and (b) centre-specific shrinkage estimates of the incremental net benefits to explore appropriately the variability between centres/countries of the cost-effectiveness results.

There are a number of ways to estimate multilevel models. Perhaps the most common methods are via maximum likelihood (assuming normality of the components of the random part of the model) and the generalised least-squares method (GLS) or an iterative version of it (which is equivalent to maximum likelihood estimation if components of the random part are normally distributed). In this chapter, models are estimated using Bayesian Markov Chain Monte Carlo (MCMC) methods. These methods are easily implemented in the package MLwiN.[223] An advantage of MCMC in this context is that it provides output in a convenient format for calculating CEACs. For a variance component model, the probability that the intervention is cost-effective is simply the probability that β_1 is greater than zero. This can be obtained in a straightforward manner from the posterior distribution of β_1. If we wish to derive the CEAC for the kth centre using a random coefficient model, then we simply derive the posterior distribution for $(\beta_1 + u_{1k})$ and again

read off the probability that this composite parameter is greater than zero for a given value of the ceiling ratio. To implement MCMC methods, we use uninformative priors and assume that the random components are normally distributed with zero mean and either constant variances (in the variance components model) or variances as a function of the predictor variables (as in the random coefficient model).

The datasets

Three different datasets are employed. Two of these are prospective economic evaluations conducted alongside multicentre trials undertaken wholly or predominantly in the UK. The third study is a cost-effectiveness analysis undertaken as part of a multinational randomised clinical trial.

EVALUATE trial

The EVALUATE trial was a multicentre randomised trial comparing laparoscopic-assisted hysterectomy and standard hysterectomy (vaginal or abdominal).[224] The study enrolled 1380 women who were first allocated by their clinician to one of the two forms of standard treatment on the basis of their clinical condition. Once assigned to one of the two standard procedures, women were randomised between laparoscopic-assisted hysterectomy and the standard procedure. The time horizon for the initial economic analysis was 1 year. Data were collected at baseline, during the main period of hospitalisation and at 6 weeks and 4 and 12 months after discharge. The present analysis uses data up to 6 weeks after discharge to avoid the problem of administrative censoring occurring at follow-up.

The study took a health service perspective and resource use measurements included length of stay in hospital, duration of surgical procedure, drug use, consumables and management of complications. In order to maximise the amount of centre-specific inpatient cost data, only patients recruited in English hospitals are included in this case study (25 out of a total of 30 centres in the UK and South Africa). Detailed hospital-specific 'hotel costs' (i.e. ward costs plus nursing) of inpatient stay were obtained from five centres. Data are available for English hospitals on the *per diem* cost of care by specialty – that is, a fully allocated cost per inpatient day.[225] Based on the mean ratio of the daily hotel cost to the full *per diem* cost in the five trial centres from which cost data were collected, it was possible to estimate hotel costs for all other English centres in the trial. National average unit costs at 1999–2000 prices were used to value all the other resource use measured in trial.[225–229] Therefore, some degree of centre specificity in unit costs has been achieved in the analysis, but it has not been possible to reflect the full variability which would be expected.

Health outcomes were expressed in terms of QALYs based on women's responses to the EQ-5D[202,230] at baseline and at each of the follow-up assessments. Given the design of the trial, in effect two separate comparisons were undertaken depending on which standard procedure patients were initially allocated. For the purposes of the present application only the comparison laparoscopic-assisted hysterectomy ($n = 573$) versus abdominal hysterectomy ($n = 286$) is included. In addition, to avoid having to deal with intermittent missing data in the present analysis, we include only observations with complete data on both costs and QALYs. Owing to these simplifications, the specific results of the analyses presented here are not the same as in the main study reports and should not be interpreted as definitive.[231]

MRC laparoscopic hernia repair trial

This was a multicentre trial conducted in 26 hospitals in the UK and Ireland, comparing open versus laparoscopic inguinal hernia repair. Details of the clinical and economic studies have been published elsewhere.[232,233] A total of 928 patients were randomised between laparoscopic ($n = 468$) and open ($n = 460$) hernia repair. The time horizon of the economic analysis was 1 year after discharge, although outcomes were calculated on a 3-month time horizon. Cost analysis was undertaken from a health service perspective. Patient-level information on resource use included time in theatre, staffing, management of complications and postoperative hospital stay. The authors identified and measured prospectively hospital-specific operative costs (equipment, consumables and sterilisation costs) from a sample of hospitals participating in the trial. These were then used to estimate operative costs in the other centres in the trial. Valuation of resource use was carried out at 1998 prices. Unit costs for consumables were obtained from the manufacturers' prices, drugs cost were estimated using the BNF[234] and staff time was costed using published statistics.[235] Outcomes were in terms of QALYs as measured by the EQ-5D at baseline, 1 month and 3 months. Because some of the observations in the original trial data set were characterised by missing data, to simplify the analysis only those patients from the trial with complete data on both cost and QALYs were included. This reduced the sample of observations

used in this application down to 547 patients. Again, owing to these simplifications, the specific results of the analyses presented here are not the same as in the main study reports and should not be interpreted as definitive.[233]

ATLAS trial

The Assessment of Treatment with Lisinopril and Survival (ATLAS) study was a multicentre and multinational trial, which enrolled 3164 patients in 19 countries, comparing low doses (n = 1596) and high doses (n = 1568) of the angiotensin-converting enzyme (ACE) inhibitor lisinopril in patients with chronic heart failure. As for the previous case studies, details of the economic and clinical analyses have been reported elsewhere.[155,236] The study adopted a health service perspective and a time horizon of 4 years. Reflecting the main anticipated cost drivers, patient-level resource use data consisted of days in hospital, day cases (including GP cost) and drugs use. Hospital costs were valued applying 1997–98 UK unit costs[237,238] to all patients' resource use. Mortality data were used to calculate expected survival duration in the two arms of the trial over 4 years of follow-up. Among the 19 countries, the USA alone recruited 38% of the entire study sample. To simplify the analysis, only costs occurring during the first year of the study are included here to avoid the problem of censored cost data. Again, owing to these simplifications, the specific results of the analyses presented here are not the same as in the main study report[155] and should not be interpreted as definitive.

Results of case studies

Summary data by centre

Tables 13–15 present summary data for the three trials by centre. These might be described as 'naïve' centre-specific results as they are generated simply by using individual centre data rather than from MLM. Columns 3–5 in each table present mean costs, outcomes (QALYs or survival) and net monetary benefits (λ = £30,000) by centre. Standard deviations are presented in parentheses. The final column of each table presents incremental *NMB*s. These were obtained by an OLS regression of *NMB* (λ = 30,000) on a

TABLE 13 *Descriptive data by centre in the EVALUATE trial: means and standard deviations (in parentheses) for costs, QALYs and* NMBs[a]

Centre	n_j	Costs	QALYs	*NMB* (λ = £30,000)	*INMB* (λ = £30,000)
1	110	1358.5 (482.6)	0.0878 (0.0194)	1276.8 (767.9)	238.4 (155.1)
2	5	1966.0 (1572.4)	0.0923 (0.0101)	802.0 (1829.0)	–1633.1 (1681.6)
3	12	1669.0 (770.8)	0.0957 (0.0135)	1202.0 (985.6)	–1014.8 (545.6)
4	71	1171.2 (440.6)	0.0833 (0.0229)	1328.0 (950.7)	296.4 (243.2)
5	26	2098.0 (676.0)	0.0952 (0.0128)	757.3 (733.3)	–775.6 (275.8)
6	11	1191.0 (253.4)	0.0785 (0.0238)	1164.7 (708.7)	658.2 (413.6)
7	13	1609.0 (328.7)	0.0983 (0.0085)	1341.0 (415.9)	–108.9 (245.4)
8	43	1757.2 (1594.3)	0.0934 (0.0179)	1044.9 (1698.9)	267.9 (558.0)
9	44	1478.2 (490.3)	0.0834 (0.0221)	1025.3 (921.1)	–117.6 (301.1)
10	72	1751.1 (1587.7)	0.0862 (0.0214)	835.7 (1924.3)	–1009.4 (469.2)
11	74	1497.6 (731.9)	0.0917 (0.0198)	1253.1 (959.1)	–6.289 (242.6)
12	56	1847.2 (701.9)	0.0885 (0.0180)	806.6 (851.5)	–385.0 (240.2)
13	3	2355.9 (218.5)	0.1095 (0.0101)	930.3 (172.2)	–76.3 (288.3)
14	8	2101.3 (1053.5)	0.0818 (0.0259)	353.5 (1496.3)	–923.0 (1118.6)
15	67	1154.2 (784.1)	0.0919 (0.0215)	1604.0 (999.3)	–279.9 (259.7)
16	52	1510.5 (671.9)	0.0891 (0.0207)	1162.9 (854.4)	–393.9 (242.2)
17	10	955.7 (137.8)	0.0855 (0.0117)	1609.4 (408.0)	248.0 (265.2)
18	32	1840.1 (716.9)	0.0931 (0.0175)	952.8 (876.3)	163.0 (330.2)
19	24	1630.4 (307.5)	0.0995 (0.0370)	1354.3 (1069.1)	269.7 (457.3)
20	72	2464.0 (1041.4)	0.0943 (0.0141)	364.2 (1150.2)	136.1 (289.1)
21	4	1754.5 (1270.2)	0.0745 (0.0178)	481.8 (1466.8)	–1155.9 (1599.7)
22	15	1575.7 (242.3)	0.0860 (0.0189)	1002.9 (634.8)	83.8 (360.1)
23	18	1658.6 (1633.6)	0.0901 (0.0044)	1045.1 (1626.1)	–620.6 (795.4)
24	7	1659.2 (455.8)	0.0830 (0.0253)	831.8 (694.4)	289.7 (811.3)
25	10	1924.5 (951.7)	0.0887 (0.0036)	737.5 (994.2)	–1166.1 (599.7)

[a] Incremental *NMB*s were calculated by regressing *NMB* on treatment arm using OLS [refer to equation (3)]. The estimated coefficients, $\hat{\beta}_1$, together with their standard errors (in parentheses), are reported. *INMB* is laparoscopic minus abdominal hysterectomy.

TABLE 14 *Descriptive data by centre in the MRC Laparoscopic Hernia Repair trial: means and standard deviations (in parentheses) for costs, QALYs and* NMBs[a]

Centre	n_j	Costs	QALYs	*NMB* (λ = £30,000)	*INMB* (λ = £30,000)
1	0	–	–	–	–
2	0	–	–	–	–
3	2	581.2 (NA)	0.1599 (0.0341)	4216.0 (1025.3)	NA[b]
4	5	1426.7 (292.62)	0.1732 (0.0293)	3770.0 (910.1)	–1099.7 (988.6)
5	0	–	–	–	–
6	29	1111.0 (389.2)	0.1673 (0.0444)	3908.0 (1413.7)	226.5 (535.8)
7	0	–	–	–	–
8	0	–	–	–	–
9	25	1142.0 (326.2)	0.1707 (0.0323)	3980.9 (1015.9)	–359.2 (408.6)
10	6	1430.9 (283.5)	0.1462 (0.0600)	2956.7 (2000.5)	–489.9 (1921.4)
11	17	1077.3 (384.8)	0.1750 (0.0362)	4173.5 (1066.9)	–87.7 (542.5)
12	2	1217.6 (300.7)	0.1962 (0.0047)	4670.3 (443.4)	–627.0 (NA)
13	9	1045.1 (442.2)	0.1487 (0.0593)	3415.9 (2085.0)	–3043.7 (955.1)
14	1	922.2 (NA)	0.2026 (NA)	5156.5 (NA)	–76.3 (288.3)
15	0	–	–	–	–
16	0	–	–	–	–
17	0	–	–	–	–
18	0	–	–	–	–
19	0	–	–	–	–
20	0	–	–	–	–
21	33	1378.9 (499.5)	0.1698 (0.0331)	3717.6 (1095.7)	–759.6 (362.9)
22	3	1282.9 (348.4)	0.2020 (0.0117)	4778.1 (655.6)	–666.9 (918.9)
23	0	–	–	–	–
24	0	–	–	–	–
25	163	616.1 (172.8)	0.1752 (0.0404)	4640.8 (1244.8)	–96.1 (195.7)
26	252	654.2 (167.5)	0.1695 (0.0439)	4430.9 (1356.0)	203.7 (170.7)

[a] Incremental *NMB*s were calculated by regressing *NMB* on treatment arm using OLS [refer to equation (3)]. The estimated coefficients, $\hat{\beta}_1$, together with their standard errors (in parentheses), are reported. *INMB* is laparoscopic versus open repair.

[b] NA, not applicable. *INMB* could not be calculated for centre 3 because the two patients recruited in this centre were randomised to the same treatment arm.

treatment dummy [refer to equation 3)]; the *INMB* is simply the estimated coefficient, $\hat{\beta}_1$. The estimated coefficient together with its standard error (in parentheses) are reported.

EVALUATE trial

In total, 859 patients were included in the study distributed across the 25 hospitals. The distribution of patients across hospitals is unbalanced, with a minimum of three patients observed in Centre 13 and a maximum 110 in Centre 1. The average number of patients per centre is 34. A simple inspection of the results reveals a great deal of variability in costs and outcomes, both within and across the centres. For example, Centre 10 has a mean cost of £1751 with a standard deviation of £1588 and mean costs range from £956 in Centre 17 to £2464 in Centre 20. This, in turn, relates to variability in *NMB*s both across and within centres. The minimum *INMB* is observed for Centre 2 (–£1633) and the maximum for Centre 6 (£658). Indeed, inspection of the results by centre shows that the sign of *INMB* also varies. Although there is appreciable uncertainty around these centre-specific estimates, they indicate that decisions based on incremental cost-effectiveness may have the potential to vary by centre.

MRC Hernia Repair trial

A total of 547 patients with complete observations on both costs and QALYs were included in this analysis. This subsample was unevenly distributed across 13 hospitals, with a minimum of one patient observed in Centre 14 and a maximum of 252 in Centre 26. The average number of patients per centre is 21. As in the EVALUATE trial, there is a great deal of variability in costs and outcomes, both within and across the centres. For example, Centre 21 has a mean cost of £1378 with a standard deviation of £499 and mean costs range from £581 in Centre 3 to £1426 in Centre 4. The

TABLE 15 *Descriptive data by centre in the ATLAS trial: means and standard deviations (in parentheses) for costs, survival and* NMBs[a]

Centre	n_j	Costs	QALYs	*NMB* (λ = £30,000)	*INMB* (λ = £30,000)
1	91	4208.8 (8686.2)	2.7413 (1.2464)	78031 (41353)	–1297 (8722)
2	25	3684.0 (6177.6)	2.4645 (1.5325)	70252 (48490)	28504 (18917)
3	108	3479.6 (6614.5)	2.9794 (1.2921)	85903 (41364)	1313 (8009)
4	290	1957.8 (4023.3)	2.8928 (1.2791)	84826 (39573)	7780 (4633)
5	148	2499.4 (5244.5)	2.8897 (1.3898)	84192 (43069)	5251 (7091)
6	100	3265.9 (8944.4)	2.9004 (1.4272)	83747 (46364)	13363 (9223)
7	52	2560.3 (3704.0)	3.0787 (1.2649)	89801 (38314)	3873 (10718)
8	125	2500.4 (4996.0)	2.9350 (1.2823)	85551 (39700)	342 (7136)
9	151	2534.0 (5326.7)	2.9901 (1.2419)	87171 (39119)	–4038 (6379)
10	74	1892.2 (3631.3)	3.0779 (1.1148)	90446 (35182)	–125 (8239)
11	72	2981.9 (8275.6)	3.1420 (1.3378)	91279 (42606)	–15703 (9938)
12	76	4024.1 (11452.2)	2.9652 (1.1989)	84934 (40505)	24073 (8930)
13	52	3573.9 (10046.9)	2.6926 (1.3035)	77206 (40845)	13187 (11296)
14	6	217.5 (202.3)	2.8228 (1.4832)	84467 (44424)	–45234 (33660)
15	72	2502.7 (5073.8)	3.4326 (1.3796)	100475 (43283)	–4933 (10261)
16	232	3191.8 (6653.7)	2.8646 (1.4095)	82748 (44362)	1012 (5837)
17	2	296.4 (248.3)	3.8863 (0.2150)	116292 (6202)	8772 (NA)
18	290	1811.7 (4160.0)	2.9719 (1.3175)	87345 (40385)	3228 (4748)
19	1198	2129.4 (5184.7)	3.0002 (1.2287)	87877 (38509)	534 (2226)

[a] Incremental *NMB*s were calculated by regressing *NMB* on treatment arm using OLS [refer to equation (3)]. The estimated coefficients, $\hat{\beta}_1$, together with their standard errors (in parentheses), are reported. *INMB* is high versus low dose.

minimum *INMB* is observed for Centre 13 (–£3043) and the maximum for Centre 6 (£226). As in the previous dataset, the sign of *INMB* varies also by centre. One additional problem with this dataset relates to the fact that 76% of the observations are clustered in two centres, with the remaining 24% distributed across 11 centres.

ATLAS

Similarly to the previous two datasets, the 3164 patients recruited in the ATLAS trial were unevenly distributed across 19 countries participating in the study. The centre with the minimum number of observations was country 17, which recruited only two patients, and country 19 was the biggest recruiter with 1198 patients. The average number of observations per country was 104. Consistent with the other two case studies, a considerable degree of **within-** and **between-**countries variability in costs and outcomes can be observed. For example, country 12 has a mean cost of £4024 with a standard deviation of £11,452. Mean costs range from £217 in Centre 14 to £4204 in Centre 12. The minimum *INMB* is observed for Centre 14 (–£45,234) and the maximum for Centre 2 (£28,504). Again, the sign of *INMB* varies also by country. Finally, 38% of the total observations are clustered in centre 19 and 73% of the total sample is distributed between six countries.

Ordinary least squares (OLS)

A comparison of the OLS results in *Tables 16–18* with those obtained from the standard analysis reported in *Tables 13–15* demonstrates the equivalence of the two approaches. In *Tables 16–18* results are expressed in terms of net benefits and are reported for three different values of λ – the value of an additional unit of benefit (i.e. $\lambda = 0$, $\lambda = 30{,}000$, and $\lambda \to \infty$).

To interpret the results from the OLS regression, it is perhaps worth revisiting the net benefit equations [equations (2c) and (2d)]. In *Tables 16–18*, the results of the OLS regression reported for the cases in which $\lambda = 0$ and $\lambda = 30{,}000$ are referring to the *NMB* [see equation (2c)], whereas the results reported for the case $\lambda \to \infty$ refer to the *NHB* [see equation (2d)]. Notice that when $\lambda = 0$, the *NMB* is equal to the negative of the cost vector. On the other hand, for the case where $\lambda \to \infty$ the *NHB* is the appropriate expression [equation (2d)] because when $\lambda \to \infty$ also the $NMB \to \infty$. In this case it is easy to verify that, when $\lambda \to \infty$, the *NHB* is equal to use the effectiveness measure as dependent variable.

It can be seen that the pooled results indicate that laparoscopic hysterectomy is associated with higher cost and higher QALYs, generating an

TABLE 16 *Results of standard economic evaluation based on pooled (OLS) analysis for the EVALUATE trial using three different specifications of net monetary benefits (*NMBs*): coefficients and 95% CI (in parentheses) are reported*

***n* = 859**	***NMB*** $\lambda = 0$	***NMB*** $\lambda = 30{,}000$	***NMB*** $\lambda \to \infty$
Constant: β_0	–1472.1 (–1581.5 to –1362.7)	1166.8 (1034.3 to 1299.3)	0.088 (0.088 to 0.092)
Treatment: β_1	–215.7 (–349.6 to –81.8)	–145.3 (–307.6 to 17.0)	0.0023 (–0.002 to 0.006)

TABLE 17 *Results of standard economic evaluation based on pooled (OLS) analysis for the MRC Hernia Repair trial using three different specifications of net monetary benefits (*NMBs*): coefficients and 95% CIs (in parentheses) are reported*

***n* = 547**	***NMB*** $\lambda = 0$	***NMB*** $\lambda = 30{,}000$	***NMB*** $\lambda \to \infty$
Constant: β_0	–649.9 (–689.4 to –610.4)	4388.5 (4227.5 to 4549.5)	0.1679 (0.16291 to 0.17298)
Treatment: β_1	–240.7 (–295.7 to 185.7)	–58.0 (–282.2 to 166.2)	0.0061 (–0.0009 to 0.01310)

TABLE 18 *Results of standard economic evaluation based on pooled (OLS) analysis for the ATLAS trial using three different specifications of net monetary benefits (*NMBs*): coefficients and 95% CIs (in parentheses) are reported*

***n* = 3164**	***NMB*** $\lambda = 0$	***NMB*** $\lambda = 30{,}000$	***NMB*** $\lambda \to \infty$
Constant: β_0	–2416.6 (–2704.0 to –2129.3)	85330.4 (83350.9 to 87309.9)	1068.3 (1045.3 to 1091.3)
Treatment: β_1	–91.1 (–499.2 to 317.1)	2311.7 (–500.2 to 5123.6)	29.2 (–3.4 to 61.9)

ICER of £93,783. At $\lambda \to 30{,}000$, this is equivalent to an *INMB* of –£145 (95% CI –£307.6 to £17.0).

Similarly, the results of the OLS regression applied to the MRC Hernia Repair (*Table 17*) data set suggest that laparoscopic hernia repair was also associated with a higher cost and higher QALY, generating an ICER of £39,460. At $\lambda = 30{,}000$, this is equivalent an *INMB* of –£58 (95% CI: –£282.2 to £166.2).

Finally, the OLS regression results applied to the ATLAS trial data set indicate that a high dose of lisinopril increased the mean cost by £91 but produced a mean survival gain of 29 days (ICER = £3.11 per day of life gained). The *INMB* at $\lambda = 30{,}000$ amounted to £2312 (95% CI: –£500.2 to £5123.6).

Multilevel analysis

The results of the multilevel analyses are reported in *Tables 19* and *20* relating to the variance components and random coefficient models, respectively. The estimation of the multilevel model specifications described in this chapter, when applied to the MRC Hernia Repair and ATLAS trials, did not converge and are not considered further in this chapter. Further investigation is required to define alternative model specifications for these two case studies. The implications of these problems is considered further in the Discussion section of this chapter (p. 64). The remainder of this chapter will concentrate on the results of the multilevel analysis of the EVALUATE trial.

Variance components models

The results of the analysis of the variance components specification for the EVALUATE trial are reported in *Table 19*. On the whole, the parameter estimates of β_0 and β_1 are similar to those obtained using OLS. However, the multilevel analysis facilitates the decomposition of the level-1 and level-2 variance components and hence the calculation of the *ICC*. This information is important as it can be employed usefully to explore the extent of the generalisability of economic evaluation results across locations. A high *ICC* indicates that the level-2 variation is an important component of the total variation and, accordingly, centres differ substantially in measured health outcomes and/or costs. In such circumstances, ignoring the hierarchical structure in the data could lead to misleading results when quantifying the sampling variability around the estimates of interest. Conversely, a near-zero *ICC*

TABLE 19 *Results of the variance components specification for the EVALUATE trial: coefficients and 95% credibility intervals (in parentheses) are reported*

***n* = 859**	***NMB*** **$\lambda = 0$** **Chain length = 50,000** **Burn-in = 2000**	***NMB*** **$\lambda = 30{,}000$** **Chain length = 10,000** **Burn-in = 1000**	***NMB*** **$\lambda \to \infty$** **Chain length = 10,000** **Burn-in = 1000**
Fixed part:			
Constant: β_0	–1509.3 (–1686.2 to –1336.4)	1142.3 (959.6 to 1316.9)	0.088 (0.082 to 0.093)
Treatment: β_1	–220.1 (–345.8 to –94.0)	–147.5 (–303.1 to 8.5)	0.0025 (–0.0003 to 0.005)
Random part:			
σ_{u0}^2	105,097 (46,894 to 210,306)	75,424 (27,612 to 166,035)	0.00015 (0.00008 to 0.0003)
σ_{ε}^2	779,731 (708,726 to 859,026)	1,234,346 (1,123,403 to 1,356,718)	0.00039 (0.00036 to 0.00043)
ICC (%)	11.90	5.80	27.80

TABLE 20 *Random coefficients specification for the EVALUATE trial: coefficients and 95% credibility intervals (in parentheses) are reported*

***n* = 859**	***NMB*** **$\lambda = 0$** **Chain length = 50,000** **Burn-in = 2000**	***NMB*** **$\lambda = 30{,}000$** **Chain length = 50,000** **Burn-in = 2000**	***NMB*** **$\lambda \to \infty$** **Chain length = 100,000** **Burn-in = 5000**
Fixed part:			
Constant: β_0	–1493.8 (–1663.3 to –1324.2)	1159.0 (986.8 to 1331.4)	0.088 (0.085 to 0.091)
Treatment: β_1	–243.8 (–407.6 to –88.1)	–176.3 (–385.6 to 23.3)	0.0024 (–0.003 to 0.008)
Random part:			
σ_{u0}^2	94,821 (35,721 to 206,039)	61,134 (15,489 to 158,570)	0.00001 (0.000002 to 0.00004)
σ_{u1}^2	40,555 (7623 to 126,824)	71,936 (14,522 to 213,021)	0.00014 (0.00007 to 0.0003)
σ_{u01}	–6430 (–72,055 to 36,071)	–13,627 (–98,753 to 32,192)	–0.000006 (–0.00003 to 0.00001)
σ_{ε}^2	772,717 (701,394 to 851,271)	1,220,511 (1,108,103 to 1,344,094)	0.0004 (0.00036 to 0.00044)
ICC			
Abdominal hysterectomy (%)	10.90	4.80	2.40
Laparoscopic hysterectomy (%)	13.70	8.00	25.70

would normally be expected to indicate that the role of level-2 variation is modest and that centres can be assumed to have similar results. The extent of clustering in these data is best observed when considering the two extreme cases of modelling cost (*NMB* with $\lambda = 0$) data and modelling health outcome (*NHB* with $\lambda \to \infty$) data. For the EVALUATE trial, the respective *ICCs* are 11.9 and 27.8%, suggesting that the degree of clustering should not be ignored.

Random coefficient models

Table 20 reports the results of estimating a random coefficient model [equation (14)]. This specification allows the effect of the intervention to vary across centres. In the EVALUATE trial, the estimated coefficients on the intervention fixed-part parameter, β_1, are similar (given the level of sampling variability evidenced by the width of the credibility intervals) to those reported for the variance coefficient model in *Table 19*.

As with the variance components specification, the random coefficients analysis confirms that the structure of the dataset is clustered, more so for cost data than health outcome data. For example, for values of $\lambda = 0$, the *ICC* was 11 and 14% in the abdominal and laparoscopic arms, respectively. The *NHB* analysis with $\lambda \to \infty$ shows considerable clustering within the laparoscopic arm but much less clustering within the abdominal arm of the trial. However, the estimates of the variance components are close to zero and the *ICC* here is acutely susceptible to sampling variability. Hence,

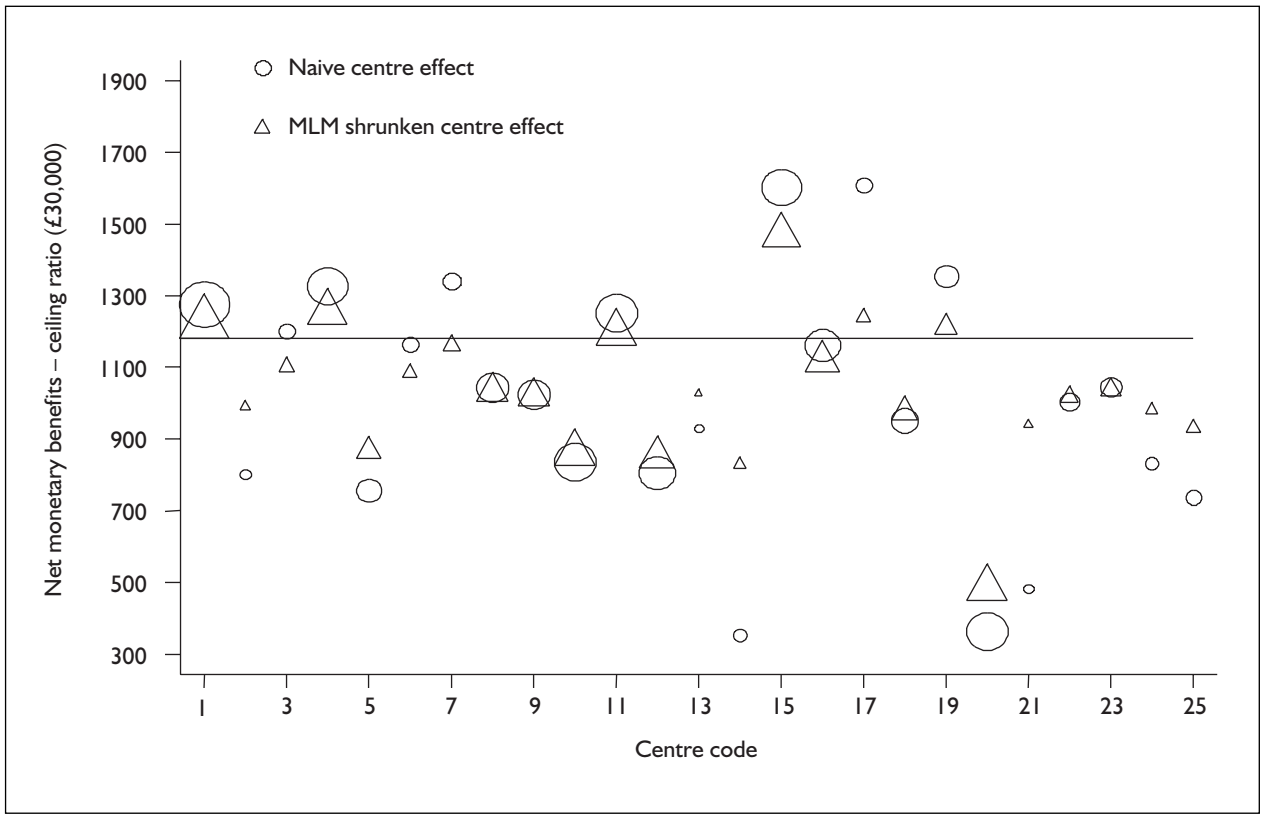

FIGURE 2 *Shrinkage estimation of centre-specific* NMBs *at* $\lambda = 30{,}000$ *based on a variance components specification (the size of the objects represents the sample size in each centre)*

although it would appear that greater correlation between patient health outcomes within centres exists for the laparoscopic arm of the trial, the absolute value of the *ICC* should be interpreted with caution.

Location-specific cost-effectiveness

Shrinkage estimation

As indicated above, location-specific estimates of cost-effectiveness are facilitated through the residual shrinkage feature of MLM. *Figure 2* illustrates the effect of shrinkage on centre-specific estimates of *NMB* [with the ceiling ratio (λ) at £30,000]. The horizontal line represents the mean *NMB* for the abdominal hysterectomy group obtained by estimating a simple variance components random effects regression model [equation (8)]; its value is simply the resulting estimated coefficient, $\hat{\beta}_0$, and for the EVALUATE trial this is £1166.80. Centre-specific *NMB*s are simply calculated as the mean *NMB* for each centre independently. These are the values given in column 5 of *Table 13* which are termed naïve centre effects and are displayed as circles.

Centre-specific shrinkage estimates are also calculated using expression equation (19). These are displayed as triangles. The relative size of the symbols is intended to provide a guide to the number of observations within each centre. It is easily seen that the naïve centre-specific estimates of mean *NMB* are always larger than the corresponding MLM shrunken estimates. Further, in general, the smaller the number of patients within a centre the greater is the discrepancy between the naïve and shrunken estimates, that is, the greater the shrinkage. This is to be expected, as the smaller the number of patients, the less information is available from which to derive the centre-specific estimate. Accordingly, one would have less confidence in the resulting estimate and this is reflected by applying greater weight to the shrinkage factor that pulls the naïve estimate towards the overall mean. The same process is applied to the calculation of centre-specific *INMBs*.

Incremental net monetary benefits

Location-specific cost-effectiveness can also be illustrated using *INMB* curves as a function of the ceiling ratio, λ (*Figure 3*). Several alternative *INMB* curves are presented. The first was obtained by specifying and estimating the single-level model [equation (3)] where, for a given value of the ceiling ratio, *INMB* was estimated as $\hat{\beta}_1$ (curve with circles). The second *INMB* curve was

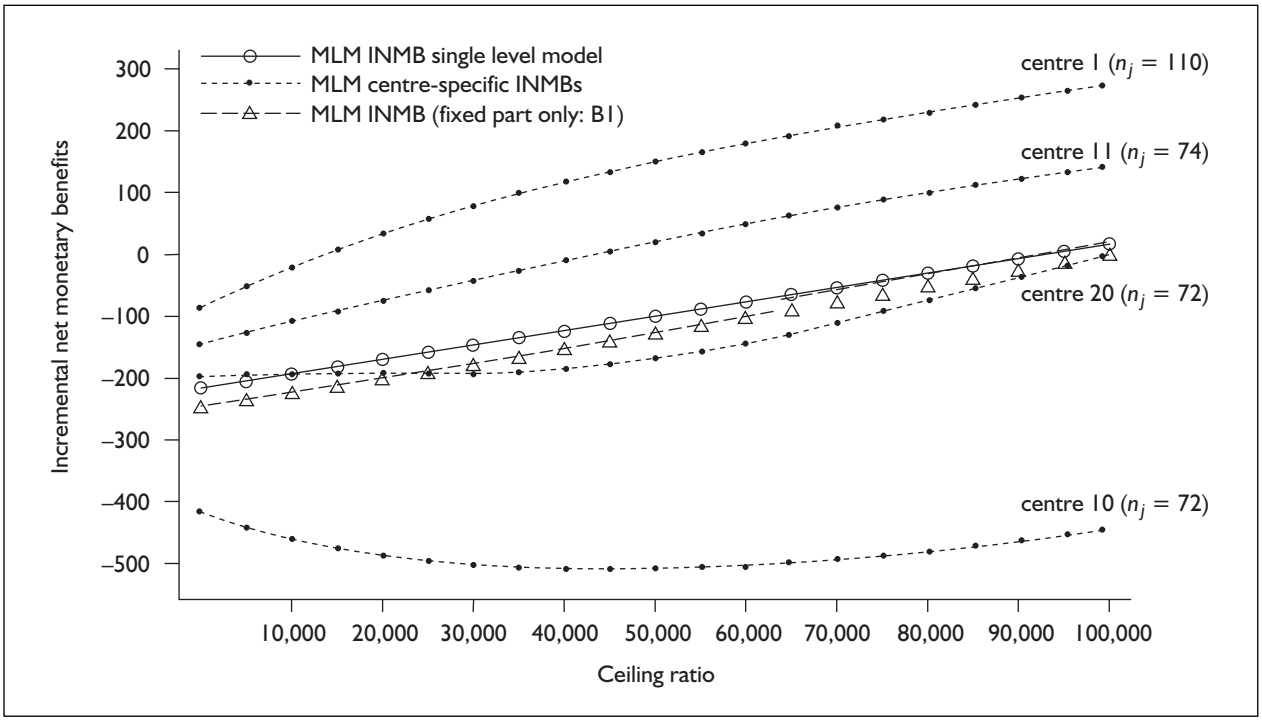

FIGURE 3 *Incremental net monetary benefit curves based on a random coefficients specification*

estimated from the **fixed part** only (again using estimate $\hat{\beta}_1$) of the random coefficient multilevel model [equation (14)] (curve with triangles). Finally, centre-specific *INMB* curves were calculated, for selected hospitals, as the sum of the **fixed** and **random** *INMB* coefficients, $\hat{\beta}_1 + \hat{u}_{1j}$, and are denoted by dotted curves. Owing to the figures becoming unwieldy when all centre-specific *INMB* curves are displayed, the largest centres are shown in the figure since these provide the most robust centre-specific estimates being based on larger numbers of patients.

A couple of points can be made about this analysis. First, the results illustrate the wide variability in hospital-specific *INMB*s. This is apparent across the full range of values of the ceiling ratio considered. This feature is also apparent for values of λ for which the corresponding model *ICC*s suggest that centre-specific variation is negligible relative to total residual variation. Indeed, the centre-specific estimates reveal that there is at least one centre for which *INMB* is positive over a wide range of values of the ceiling ratio. This is evident for Centre 1, for which the results are contrary to those obtained by considering only the fixed-part estimate of incremental cost effectiveness $\hat{\beta}_1$, obtained through either the single-level specification or the random coefficient specification, both of which display negative estimates for values of λ up to £95,000. Second, the centre-specific curves are a non-linear function of the ceiling ratio. This is due to the non-linear relationship between the ICC and the ceiling ratio. This affects the random part estimate of the *INMB* ($\hat{u}_{1j}$) through the shrinkage estimation process since this is, in turn, a function of the *ICC*.

Cost-effectiveness acceptability curves

Similar features are revealed in terms of CEACs (*Figure 4*). Again, circles denote the curve produced from a single-level model specification [equation (3)], while triangles denote the curve obtained from the fixed part only (using estimate $\hat{\beta}_1$) of the random coefficient multilevel model [equation (14)]. Curves are also shown for the same hospitals used in the *INMB* analysis in *Figure 3*.

Once again, the curves display great variability across hospitals in cost-effectiveness for given values of λ. This variability appears greatest at the value of λ of around £60,000, although caution is required here as this observation is based on only those selected hospitals displayed. For example, the probability of laparoscopic hysterectomy being cost-effective, with a ceiling ratio of £50,000, is approximately 0.16 applying the results of the single-level model or fixed-part random coefficient model specification. The corresponding probability for Centre 10 is $p = 0.01$ and for

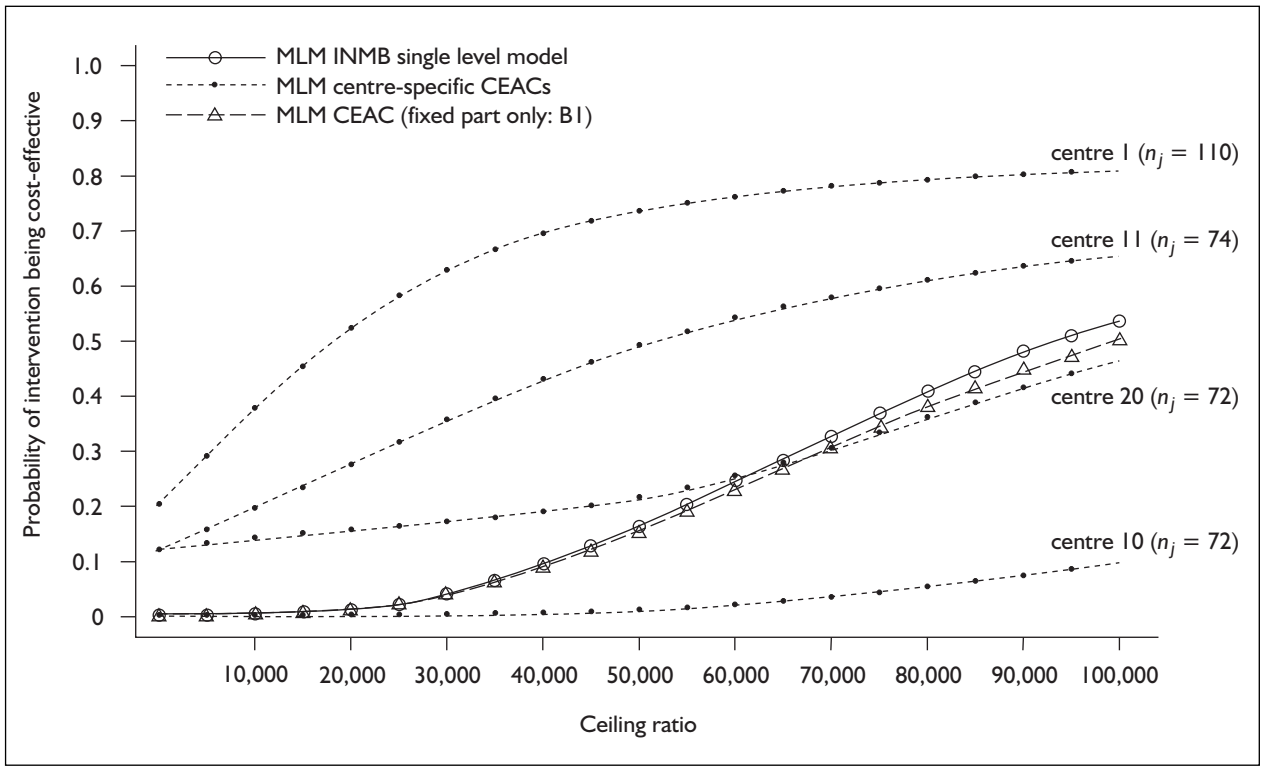

FIGURE 4 *Cost-effectiveness acceptability curves based on a random coefficient specification*

Centre 1 $p = 0.72$. The observed maximum probability that the intervention is cost-effective for Centre 10 is $p \approx 0.1$ (at $\lambda = 100{,}000$). For Centre 1, the maximum is $p = 0.8$. For values of $\lambda >$ £20,000, laparoscopic hysterectomy would probably be considered cost-effective based on the results of Centre 1. However, for values of λ less than at least £100,000, laparoscopic hysterectomy would probably not be considered cost-effective based on patient cost and outcomes reported for Centre 10. Results obtained from the single-level model specification indicate that the probability that the intervention is cost-effective is less than $p = 5$ for the values of λ considered.

Discussion

The use of MLM is a natural extension to the regression framework advocated by Hoch and colleagues.[156] The approach formally segments variation in net benefit into that occurring at the level of individual patients and that occurring at the level(s) of locations such as centre and country. Where costs and/or outcomes data are clustered by location, these methods facilitate correct estimates of the uncertainty in cost-effectiveness results. MLM also provides a means of estimating location-specific cost-effectiveness. Here 'location' could mean centres (e.g. hospitals) within a jurisdiction or countries. In the past, this has been attempted in various ways, including the application of location-specific unit costs to all resource use and outcome data and analysing only data relating to patients randomised in the location of interest (see Chapter 5). More recent methods have used multiple regression and modelled location, and its interaction with treatment, as fixed effects.[25] The MLM framework presented here, however, uses MCMC estimation methods to predict location-specific estimates of cost-effectiveness. As part of this, location-specific estimates are shrunk back to the overall mean, with the degree of shrinkage depending on the sample size in a centre and the level of variability in net benefits between and within centres. In other words, using these methods, a location-specific estimate of cost-effectiveness is a combination of the overall mean across patients in all location, and the mean in patients within that location.

The extent to which the use of MLM is crucial in a particular study depends on the proportion of overall variability in net benefit that takes place between locations. Although, with relatively small values of the *ICC*, reasonably good agreement between the multilevel and the OLS estimates can be expected, in practical terms it is impossible to establish a rigid threshold value of the *ICC* above

which the use of MLM should be recommended. Given this, there is an argument for starting any CEA alongside a multilocation trial with a multilevel regression to understand the data structure in more depth.

It is recognised that, if MLM is to be used routinely for CEA, additional methods research will be required. In particular, the results of the MLM estimation applied to the MRC Hernia Repair and ATLAS case studies have not been presented because of problems in identifying an appropriate model specification and further analysis of these types of datasets is necessary.

The clustering structure of the dependent variable (net benefit) is affected by the structures of costs and outcome data. For larger values of λ, the outcome data (e.g. QALYs) will be progressively more important in the net benefit expression. In the EVALUATE trial here, it was found that cost data tended to be more clustered by location than outcome data. Consequently, for λ values close to zero where the clustering structure of the net benefits is mainly affected by the cost data, MLM is particularly important to analyse these data. On the other hand, when λ values tend to be large and the structure of the net benefit data is influenced mainly by outcomes, MLM is perhaps less important for these datasets. However, it would be premature to generalise from the case studies here to all economic evaluations. Further research is planned using MLM to analyse patient-level datasets with longer follow-up in costs and outcomes and in disease areas where QoL weights may be more different between the treatment groups.

An important policy issue is raised by this work, that is, the extent to which location-specific estimates of incremental net benefit are useful to decision-makers. In the context of multinational trials, the ability to generate estimates of cost-effectiveness by country would seem potentially useful to country-level decision-makers. In the case of multicentre trials in a single country, however, this may not be so straightforward. The implication of generating centre-specific estimates is that the decision-maker may be willing to fund a particular intervention in one centre, but not another. Alternatively, the analysis might be useful to centre-level decision-makers. The use of location-specific covariates in MLM to explain variation in cost-effectiveness between centres and countries may have more direct policy relevance. For example, an intervention might be cost-effective in centres with high levels of patient throughput, but not in centres treating fewer patients. This might provoke policy makers to reassess the appropriate organisation of services. Although the implications of this need further consideration, it remains the case that the multilevel population average from a random effects model (e.g. see *Figure 4*) remains the most appropriate way of estimating average cost-effectiveness (i.e. across centres), given clustered data, if the decision-maker is not interested in implementing different decisions in different centres.

The analyses presented here can be extended in two important ways. The first would be to include location-specific unit cost data to value resource use measured in that location. Given the difficulty in acquiring these data in all locations, the case studies presented here use fixed unit costs for most resources. This has the effect of constraining the degree of variation between locations as a proportion of this will be generated by differences in unit costs. Furthermore, as described in Chapter 3, economic theory would suggest that unit costs and resource use will be related by location-specific production functions.[119] For example, if medical staff are very costly in one location (perhaps due to shortages), there may be a substitution of medical inputs by other professionals with different costs. In future research where it is possible to identify location-specific unit costs, MLM can be used to explore this type of substitution more formally.

The second extension to the MLM framework presented here is the use of additional covariates within the regression analysis. For ease of exposition, the only covariate included in the methods described here [equation (14)] is the treatment arm to which a patient was randomised. The inclusion of additional covariates at hospital and patient level may help to reduce the estimated variability in the model. The net benefit regression approach outlined by Hoch and colleagues[156] describes the value of including patient-level variables in these models (e.g. baseline clinical variables, socio-demographic characteristics). Their inclusion offers a way of explaining a greater proportion of the overall variation in net benefit (i.e. more precise estimates of incremental net benefit). Through the use of interactions between these level-1 covariates and treatment, it is also possible to undertake subgroup analysis, that is, to estimate incremental net benefit in subgroups of patients (e.g. defined by age and/or gender).

Within the framework of MLM, it is also possible to include covariates which explain some of the variation between the higher level groups in the model (e.g. hospitals or countries). These covariates might include factors such as patient throughput and experience of clinical staff. The rationale is the same as for patient-level covariates: their inclusion can allow more efficient estimates of cost-effectiveness, and it is possible to explore cost-effectiveness in subgroups of locations. Further research is required to identify higher level covariates which are empirically useful in these models, theoretically well founded in terms of explaining differences in efficiency between locations and useful for policy-making purposes.

In the case of multinational trials, MLM facilitates the prediction of country-specific costs and effects, which can then be used to generate better parameter estimates for decision models.

Conclusions

This chapter has introduced an analytical framework – MLM – which has potential to facilitate estimates of cost-effectiveness which both reflect the variation in costs and outcomes between locations and also enable the consistency of cost-effectiveness estimates between locations to be assessed directly. Through the use of data from three RCTs as case studies, the chapter has shown the importance of establishing, through the *ICC*, the proportion of overall variability in cost-effectiveness (net benefit) which is between locations. It has also illustrated how the uncertainty in cost-effectiveness results will be potentially misleading if standard methods are used with datasets with clustered data. Finally, the value of shrinkage estimators using MCMC estimates location-specific cost-effectiveness has been shown.

Chapter 7

Assessing generalisability in model-based economic evaluation studies: a structured review in osteoporosis

Decision analytic modelling is widely used in health technology assessment to evaluate the effectiveness and cost-effectiveness of alternative options under conditions of uncertainty.[239,240] This form of modelling is necessary in situations where a single primary source of data (e.g. a randomised trial) does not wholly satisfy the data needs of a decision problem and additional data sources and assumptions are needed. These situations include synthesising information when a number of estimates of a particular parameter exist, extending the results of a short-term trial over a long-term time horizon, increasing the range of alternative options being compared and adapting the results of a study undertaken in one location to be relevant to another jurisdiction.

In the context of this report's focus on issues of variability and generalisability in cost-effectiveness between locations, the use of decision models is of interest for two reasons. First, models are often used to take the results of a particular patient-level study, such as a randomised trial, which was undertaken in one location, and make them relevant to one or more alternative locations. Chapter 4 reviewed several studies of this type and Chapter 8 describes a modelling case study with the objective of adjusting the results of trials in unstable angina to a UK context.

The second area of interest in model-based economic studies is to assess how geographical variability is dealt with in published modelling studies. The general purpose of models is to identify optimal solutions to specific policy problems as faced by particular decision-makers or jurisdictions. For example, NICE's technology appraisal process often includes decision models submitted by manufacturers and/or developed by the academic Technology Assessment Team (see www.nice.org.uk). In each case, the decision-maker (the NICE Appraisal Committee) and the jurisdiction (NHS) is explicit. Therefore, it is of value to review published modelling studies to assess how clearly the authors have identified the decision-making 'target' of their work, the extent to which the data incorporated into these models are the most appropriate for that decision-maker, the degree to which they have assessed the importance of any variability between locations **within** a particular jurisdiction and the generalisability of their results to other decision-makers.

Rather than select a sample of model-based studies across a range of diseases and interventions, the particular clinical area of osteoporosis has been selected for the review. This clinical area has been selected for two important reasons. First, the area has a large number of cost-effectiveness models – indeed, the large majority of cost-effectiveness studies in osteoporosis are model-based. Second, there is likely to be pronounced variability between countries and other locations in many of the parameters going into these models. This variability potentially relates to baseline event rates, resource use, cost and utility.

The advantage of focusing on one clinical area is that it is possible to remove some of the variation between models which would be expected when comparing analyses in different diseases and which is not really relevant to the purpose of this report (e.g. different outcomes, treatment options, sources of uncertainty). In a review of economic evaluations in this field undertaken in 1998, a total of 16 studies were identified, all of which were based on decision analytic models.[241] Osteoporosis is characterised by low bone mineral density (BMD) and deterioration of bone tissue, leading to enhanced bone fragility and consequent increase in fracture risk. The disease is manifested in terms of a high occurrence of hip, wrist and vertebral fractures and is most prominent in postmenopausal women. Development of fractures is a complex function of osteoporosis, age and other risk factors which evolve over time. The majority of trials in this area have evaluated the impact of osteoporosis treatments on the intermediate end-point of BMD.[242] Historically, there has been a scarcity of literature evaluating

final end-points.[243] The need to estimate links between intermediate end-points and ultimate measures of health gain, together with the requirement to synthesise results from several studies and extrapolate over a long-term time horizon, are the main reasons for the preponderance of decision models in osteoporosis.

Methods

Aims and objectives

The aim of the review was to assess how published model-based economic evaluations in the field of osteoporosis deal with variability in results by location.

The systematic review set out to address the following specific objectives:

1. To assess whether the study clearly defines the decision-making audience for the model.
2. To establish whether the model was transparently reported in terms of study question, structure and data inputs.
3. To assess the relevance of the data inputs used in the model to the stated decision-maker or jurisdiction.
4. To assess how fully the robustness of the model's results to variation in data inputs between locations was assessed.

Inclusion and exclusion criteria

The inclusion criteria were full economic evaluation models evaluating therapeutic interventions in osteoporosis. Only studies reporting a summary measure of cost-effectiveness (e.g. cost-effectiveness ratio) were included in the review, as these studies combine an estimate of both costs and effectiveness and present an overall assessment of the value for money of the alternative interventions of interest to decision-makers. Since this was a review where methodology rather than results was of primary interest, only studies published in English were included. Economic evaluations which did not describe the structure and assumptions underlying a model were excluded from the assessment, as were simple cost analyses, secondary reviews of economic evaluation models and studies that did not present a summary measure of cost-effectiveness.

Search strategy

A search strategy was devised to retrieve published papers reporting economic evaluation models of interventions for osteoporosis. The searches were conducted using the following bibliographic databases: MEDLINE, EMBASE, EconLit, HEED, the internal catalogue of the CRD/CHE information service, the Health Technology Assessment (HTA) Database and the administration version of the NHS EED held at the CRD. The databases were searched from 1980 onwards, or from the earliest publication in the relevant database after that date, and searches were restricted to English-language documents. All the references from the database searches were imported into an Endnote Library and de-duplicated. The search strategies are detailed in Appendix 4. Searches were extended to bibliographies of retrieved articles, and reference lists of key review articles in the area were scrutinised. Papers that appeared relevant to the review were retrieved and assessed according to the inclusion criteria.

Review process

A data extraction tool was developed specifically for the purpose of this review. For included studies, information relevant to the review was extracted into a data extraction form by one of the authors (HU). The information was summarised in data tables, which provided the basis for assessment of the studies. Key characteristics of the models presented in each publication were recorded, as was information on interventions studied and results of the evaluations. The four numbered sections of the data extraction form addressed the four research questions of the review, and further details are provided below.

Definition of target decision-maker or jurisdiction

Being aware of the target decision-making audience for a model is important to a judgement about the appropriateness of the model and its inputs. An attempt was therefore made to elicit the target decision-making audience or jurisdiction of the model. In some instances, where models did not explicitly state the target decision-maker, it was possible to infer a decision-maker from the perspective taken and the data incorporated. Related to this, the studies were assessed according to whether the study question was clearly stated or not or whether the study question could be easily inferred.

Transparent reporting of model specification

Transparent reporting of a model is a prerequisite to understanding the relevance of the model to the target decision-maker and also to assessing its generalisability to other decision-makers and jurisdictions. The specification of study setting

(e.g. country and primary or secondary care) and patient population was therefore extracted. In addition, the description and justification of alternative interventions were considered and an assessment of transparent reporting of the model structure and key assumptions was made.

Relevance of data inputs to target decision-maker or jurisdiction

The ease with which model inputs can be traced and the relevance of those inputs to the stated decision-maker will influence the degree to which a model is considered applicable in the target setting. The reporting of sources and the relevance of key data inputs to the model were therefore assessed, covering clinical data and their valuation, resource use and unit costs. Models that reported and referenced both baseline risks and risk reductions were considered as having provided sources of clinical data, whereas models that only reported references for risk reduction were assessed as having partially provided sources of clinical data.

Assessment of robustness of model to variation in data inputs within and between jurisdictions

The use of sensitivity analysis to explore the robustness of model results to variation in data inputs that may exist within and between jurisdictions was assessed. As discussed in earlier chapters, resource use estimates and their valuation, and also health state estimates and their valuation, may vary across settings. This variation may exist between units **within** a given decision-maker's jurisdiction and, to provide information to the decision-maker, the implications of this variation should be assessed. The authors may also choose to assess the robustness of their model's results to the level of variation that might be expected **between** jurisdictions. For example, the average length of stay in hospital for patients with hip fracture has been reported to be 29.6 days in Aberdeen and 41.7 days in Peterborough,[244] whereas the national average in Denmark has been reported to be 21 days.[245]

Similarly, the estimate of clinical effect incorporated into a model may depend on the study population and target population of the model. The robustness of the model's results under a range of clinical effectiveness estimates from different studies could, therefore, be explored in sensitivity analysis. Models may also undertake adjustments to translate the results of explanatory trials that may not hold in routine clinical practice. For example, as discussed in Chapter 3, compliance is generally acknowledged to be higher within the context of randomised trials than in clinical practice, and this may contribute substantially to the reduction in efficacy when an intervention is used in a non-trial environment.[40,246] Reduced compliance in a clinical practice setting may result in reduced effectiveness of the drug, so models that evaluate the population-based impact of a strategy in clinical practice may provide a more representative estimate by factoring in the reduced compliance in the analysis.[44]

Finally, a straightforward check can be made for generalisability of a model's results, both within and between jurisdictions, by comparing the results with those of related studies reported in the literature focusing on the same or other jurisdictions.

Results

Summary of included papers

A total of 18 publications reporting economic evaluation models satisfied the inclusion criteria (*Table 21*). These included four Markov state transition models[243,247–249] and four simple decision trees.[250–253] Nine studies were CUAs. Of the studies published in the 1980s, six were from the USA. Eight studies were from European countries, all of which were published in the 1990s. The studies covered the following countries: Australia, Canada, Denmark, UK, USA and Italy. Apart from the Italian study,[253] all results were presented in local currencies. Many models based the evaluation on a time horizon spanning the lifetime from the onset of treatment at menopause, usually assumed to be 45 or 50 years of age.[243,245,248,249,254–259] Others evaluated a more limited time horizon, for example 2 years in Francis and colleagues,[250] 3 years in Rosner and colleagues,[260] 3–4 years in Torgerson and Kanis[252] and 1 year in Visentin and colleagues.[253]

Overall, the main base-case results of the studies did not reveal any systematic differences across the studies that might be explained by location (*Table 21*). Neither systematic variation in cost-effectiveness estimates within countries or systematic changes over time were apparent from the review. Despite focusing on the osteoporosis area, the models compared a range of interventions and presented results using a variety of outcomes. For example, Ankjaer-Jensen and Johnell[245] reported average cost per hip fracture avoided in screened and unscreened populations, comparing three different interventions. In contrast, Tosteson and colleagues[248] compared

TABLE 21 *Studies included in the review and the base-case results of the evaluations*

Study (year)	Interventions evaluated	Incremental cost-effectiveness ratio	Incremental cost per QALY
Ankjaer-Jensen, 1996[245]	Screening or not followed by calcium supplementation, etidronate, calcitonin or HRT	Cost per hip fracture avoided: population based vs screening: Calcitonin: 1,136,200 DKK Etidronate: 77,567 DKK Calcium: 522,300 DKK	NA
Cheung, 1992[254]	Oestrogen replacement therapy alone vs no treatment and oestrogen and progestin vs no treatment	NA	10 years of treatment ORT: US$9500–17,500, symptomatic women ORT: US$45,800, dominated, asymptomatic women ORT: US$6510–1,020,000, hysterectomy CRT: US$26,100–1,450,000, asymptomatic women CRT: US$9820–34,700, symptomatic women
Coyle, 2001[261]	Calcitonin was compared with alendronate and etidronate	Cost per life-year gained. 5-year treatment of 65 year old: Alendronate dominant strategy ICER for nasal calcitonin vs no therapy: Can$63,500 ICER for nasal calcitonin vs etodronate: Can$42,300	5-year treatment in 65 year old: Alendronate dominant strategy ICER for nasal calcitonin vs no therapy: Can$46,500 ICER for nasal calcitonin vs etodronate: Can$32,600
Daly, 1992[255]	Oestrogen only vs no treatment (no uterus) and oestrogen vs oestrogen + progestin (uterus)	Cost per life-year gained (10 years of treatment): ORT: £2900 in women with no uterus ORT: £8300 in women with uterus CRT: £14,400 in women with uterus	10 years of treatment in mildly symptomatic women ORT: £1700 in women with no uterus ORT: £4400 in women with uterus CRT: £6200 in women with uterus
Daly, 1996[256]	Oestrogen only vs no treatment (no uterus) and oestrogen vs oestrogen + progestin (uterus) vs no treatment	NA	Mildly symptomatic women ORT: £310 (5 years of treatment) to £660 (20 years of treatment) CRT: £550 (5 years of treatment) to £1250 (20 years of treatment)
Francis, 1995[250]	Calcitonin vs no treatment, estrogen + progestin vs no treatment, etidronate vs no treatment	Cost per vertebral fracture avoided, treatment vs no therapy: HRT: £138–680 Etidronate: £1880 Calcitonin: £9075–25,013	NA
Garton, 1997[257]	Interventions (screening vs no screening) clear but not justified A strategy of oestrogen + progestin with screening vs a strategy of oestrogen + progestin without screening (universal)	Cost per fracture prevented HRT vs no therapy: Screening: £1000–4200 Universal treatment: £1200	NA
Geelhoed, 1994[247]	Oestrogen vs no treatment and lifestyle regime (exercise) vs no treatment	NA	Intervention vs no treatment: ORT for life from 50 years: US$8810 ORT for 15 years: US$16,500 ORT from age 65 years: US$8450 Calcium and exercise: US$28,480

continued

TABLE 21 *Studies included in the review and the base-case results of the evaluations (cont'd)*

Study (year)	Interventions evaluated	Incremental cost-effectiveness ratio	Incremental cost per QALY
OTA, 1995[243]	Estrogen (ERT) vs estrogen + progestin (PERT) following different screening + treatment strategies	Mean cost per life-year gained ORT (screening): US$22,431–151,392 depending on duration of therapy and screening threshold ORT (population-based): US$23,334–126,876 depending on duration of therapy. Life-long therapy most cost-effective	NA
Rosner, 1998[260]	Estrogen, estrogen + progestin, alendronate, etidronate, calcium combinations and no intervention	ICER: cost per averted vertebral fracture for four non-dominated strategies in order: (i) calcium – no therapy (ii) HRT – calcium – no therapy: ICER = Can$166 (iii) HRT – etidronate – calcium – no therapy: ICER = Can$2331 (vi) HRT – alendronate – calcium – no therapy: ICER = Can$40,965	NA
Torgerson, 1993[251]	Compares a strategy of estrogen only in high-risk patients (screening) vs estrogen + progestin in population-based treatment	Cost per averted hip fracture Universal treatment for 10 years: £40,080 Screening, then treatment: £34,971	NA
Torgerson, 1995[252]	Vitamin D and calcium vs no treatment and vitamin D vs no treatment	Cost per averted hip fracture Oral vitamin D + calcium: £17,379 in community (low BMI); £4735 in nursing homes; cost saving overall in nursing homes (low BMI) Vitamin D injection: all options cost saving overall	NA
Tosteson, 1990[248]	Screening vs no screening No intervention, screening followed by HRT (estrogen + progestin) and universal HRT (estrogen + progestin)	NA	Screening and then treatment (15 years): US$4200–37,800 depending on treatment threshold Universal treatment (15 years): US$144,000
Tosteson, 1991[249]	Estrogen vs no intervention (intact uteri), estrogen + progestin vs no intervention (hysterectomy)	NA	ORT (in women without uterus): US$7010–9020; without QoL side-effects: US$9930–14,940 CRT (in women with uterus): US$32,660–33,780; without QoL side-effects: >US$150,000
Visentin, 1997[253]	Calcitonin vs no treatment	Cost per averted hip fracture: Calcitonin vs no treatment: Population-based approach: US$2,367,987 Screening and treatment of high-risk patients: US$838,120	NA
Weinstein, 1980[258]	Estrogen vs no treatment	NA	ORT in women with uterus: US$7420–18,160 ORT in women with osteoporosis: US$5460–15,100

continued

TABLE 21 *Studies included in the review and the base-case results of the evaluations (cont'd)*

Study (year)	Interventions evaluated	Incremental cost-effectiveness ratio	Incremental cost per QALY
Weinstein, 1983[259]	Oestrogen vs no treatment and oestrogen + progestin vs no treatment	NA	ORT: US$130,000 vs CRT CRT: US$42,000 (5 years of treatment) to US$24,000 (10 years of treatment) vs no treatment
Weinstein, 1990[262]	Oestrogen vs no treatment and oestrogen + progestin vs no treatment.	NA	ORT for 5 years: US$72,100 in asymptomatic women ORT for 5 years: US$12,600–33,100 in symptomatic women depending on symptom relief CRT for 5 years dominates ORT for 5 years assuming QoL differences are minimal CRT for 15 years: US$22,650

BMI, body mass index; CRT, oestrogen only therapy; HRT, ORT, oestrogen and progestin therapy; NA, not applicable.

costs and QALYs of two different interventions in patient populations with different life expectancy.

Target decision-maker

None of the studies evaluated cost-effectiveness from a broader perspective than that of the health service sector. The target decision-making audience was only explicitly stated in three (16%) of the included studies (*Table 22*). Specifically, Coyle and colleagues stated the decision maker to be "a Canadian provincial Ministry of Health",[261] OTA stated that the report was commissioned by the US Senate Special Committee on Aging[243] and Visentin and colleagues commented that the evaluation was targeting the Italian Health Service.[253]

It was possible, however, to infer a target decision-maker for many of the remaining studies. For example, Ankjaer-Jensen and Johnell commented that the analysis was "carried out in a Danish context",[245] and the study by Cheung and Wren seemed to target a decision-making body in New South Wales, Australia.[254] Similarly, Daly and colleagues[255] and Torgerson and Reid[251] appeared to tailor their analyses to be applicable in British context without explicitly stating this. However, it was not possible to infer a decision-making audience in eight studies.

The research question was explicitly stated in seven studies (39%) (*Table 22*). For example, Coyle and colleagues expressed the economic study question as "to assess the cost-effectiveness of nasal calcitonin compared with no therapy, alendronate or etidronate in the treatment of postmenopausal women with previous osteoporotic fracture".[261] It was possible to infer the study question for those studies that explicitly stated the strategies under comparison and outcome at the outset. For example, Torgerson and Reid did not explicitly state a research question but compared the (average) cost-effectiveness of screening followed by hormone replacement therapy (HRT) treatment versus no screening and universal treatment.[251]

Transparency of reporting

Several aspects of transparency of reporting the models were explored (*Table 23*). Most of the studies specified that the target population was women living in the community, though one study specifically considered women in nursing homes,[251] and the target country could be inferred for all studies. The target populations of all models were also indicated. Gender, postmenopausal status or age above 50 years were used in most studies to identify the study population. Susceptibility to osteoporosis, either via previous fracture[261] or low BMD,[252] were also used as descriptive factors. Some models evaluated hysterectomised women separately from those with uterus *in situ*.[247,251,255,256,259] Two studies restricted the analysis to apply to Caucasian women only[247,250] owing to an underlying difference in baseline hip fracture risk.

The model structure was described in most of the studies through an outline of all clinical outcomes incorporated into the structure or, where relevant,

TABLE 22 *Target decision-maker and study question*

Study (year)	Q1.1 Decision-maker	Q1.2 Economic study question
Ankjaer-Jensen, 1996[245]	Inferred "The analysis was carried out in a Danish context"	Inferred Cost-effectiveness of population-based treatment programme compared with that of screening programme
Cheung, 1992[254]	Inferred Study was carried out in New South Wales, Australia	Explicit
Coyle, 2001[261]	Explicit The target decision-maker of the model was clearly stated as "a Canadian provincial ministry of health"	Explicit The economic study question was clearly outlined as "to assess the cost-effectiveness of nasal calcitonin compared with no therapy, alendronate or etidronate in the treatment of postmenopausal women with previous osteoporotic fracture"
Daly, 1992[255]	Inferred The authors state that "an assessment … in the British context is timely"	Inferred Three explicit treatment strategies were compared
Daly, 1996[256]	Inferred "Costs to the NHS …"	Explicit
Francis, 1995[250]	No	No
Garton, 1997[257]	No	Inferred Bone mass measurement compared with no measurement (i.e. screening)
Geelhoed, 1994[247]	Inferred Western Australian health service	Inferred Four strategies were compared
OTA, 1995[243]	Yes Report commissioned by the US Senate Special Committee on Aging	Yes Strategies of screening and HRT treatment compared
Rosner, 1998[260]	Inferred Clinical decision-makers in Canada	Yes Four explicit strategies compared
Torgerson, 1993[251]	Inferred The NHS in the UK	Inferred Screening vs no screening followed by HRT
Torgerson, 1995[252]	No	Yes Estimate cost per averted fracture for two interventions
Tosteson, 1990[248]	Inferred	Yes
Tosteson, 1991[249]	Inferred	Inferred
Visentin, 1997[253]	Yes Italian Health Service	Inferred
Weinstein, 1980[258]	No	Inferred
Weinstein, 1983[259]	No	Yes "To synthesize the available evidence in comparing the costs, risks and benefits of estrogen–progestin therapy and estrogen-only therapy in postmenopausal women"
Weinstein, 1990[262]	No	Inferred

TABLE 23 *Transparency of reporting of the model*

Study (year)	Q2.1 Study setting specified?	Q2.2 Patient population specified?	Q2.3 Alternative interventions stated and justified?	Q2.4 Model structure clearly stated?	Q2.5 Main assumptions clearly stated and justified?
Ankjaer-Jensen, 1996[245]	Yes	Yes	Yes	Yes	Yes
Cheung, 1992[254]	Yes	Yes.	Yes	Yes	Yes
Coyle, 2001[261]	Yes	Yes	Yes	Yes	Yes
Daly, 1992[255]	Yes	Yes	Yes	Yes	Yes
Daly, 1996[256]	Yes	Yes	Yes	Yes	Yes
Francis, 1995[250]	Yes	Yes	Yes	Yes	Yes
Garton, 1997[257]	Yes	Yes	Yes.	Yes	Yes
Geelhoed, 1994[247]	Yes	Yes	Yes	Yes.	Yes
OTA, 1995[243]	Yes	Yes	Yes	Yes	Yes
Rosner, 1998[260]	Yes	Yes	Yes	Yes	Yes
Torgerson, 1993[251]	No	Yes.	Yes.	Yes	Yes
Torgerson, 1995[252]	Yes	Yes	Yes	Yes	Yes
Tosteson, 1990[248]	Yes	Yes	Yes	Yes	Yes
Tosteson, 1991[249]	Yes	Yes	Yes	Yes	Yes
Visentin, 1997[253]	Yes	Yes	No	No	No
Weinstein, 1980[258]	Yes	Yes	Yes	Yes	Yes
Weinstein, 1983[259]	Yes	Yes	Yes	Yes	Yes
Weinstein, 1990[262]	Yes	Yes	Yes	Yes	Yes

health states and transitions. One study did not present the structure adequately:[253] it was, for example, unclear how this model estimated hip fracture risk reduction.

The main assumptions were clearly stated and justified in most studies (*Table 23*). For example, Coyle and colleagues,[261] OTA[243] and Tosteson and Weinstein[249] clearly presented all assumptions in the model as well as omissions from the models. The study by Visentin and colleagues[253] was not entirely transparent – for example, it was not clear which trials were used to populate the model.

Model input and relevance to stated decision-maker

Clinical data and health state valuation

A common feature of the models in this review was the use of epidemiological studies to estimate the relative hip fracture risk reduction in the treated populations. Three models based the hip fracture risk reduction estimates on individual clinical trial data,[250,252,260] whereas two models based the effect estimates for one of the therapies on meta-analyses of several trials.[245,261] Of these studies, only two[250,261] provided information about the clinical characteristics of the population of the trials (*Table 24*).

The remaining studies based the effect estimate primarily on observational studies. One of these studies[243] provided details of patient characteristics in the studies on which the effectiveness estimate was based and two further studies presented patient characteristics only in terms of age range.[248,260]

Since the majority of studies provided only limited patient information, the scope for assessing the applicability of the results to the target population – and indeed other populations – was constrained. Only two studies failed to provide any basis for the assumption on clinical effect.[251,254] In spite of the fact that limited information was provided on the population sample, the majority of the studies provided references to primary studies that are likely to have given a more comprehensive description of the relevant patient sample. Only one study adjusted the risks measured in the trials to the target population for the modelling exercise.[260] In spite of this, most studies appeared to use the best available data relevant to the stated or inferred decision-maker (*Table 25*).

Of the nine cost–utility studies that were included in the review, only two used utilities derived from patients.[256,261] Rosner and colleagues[260] used a Canadian Delphi panel and the Health Utilities Index as a basis for the utilities used in their model. All the other authors based their utility weights on those assumed by Weinstein[258] either implicitly or explicitly (*Table 24*). With the exception of the three studies mentioned above which used sample-based utility weights, it was difficult to assess the relevance of the utilities assumed by Weinstein[258] to particular healthcare decision-makers or jurisdictions (*Table 25*).

TABLE 24 *Information on sources for the model inputs*

Study (year)	Q3.1a Sources of clinical data provided?	Q3.2b Sources of resource use provided?	Q3.1c Sources of unit costs provided?	Q3.1d Sources of preferences/ utilities provided?
Ankjaer-Jensen, 1996[245]	Yes	Yes	Yes	NA
Cheung, 1992[254]	No	Yes	Yes	Yes
Coyle, 2001[261]	Yes	Yes	Yes	Yes
Daly, 1992[255]	Yes	Yes	Yes	Yes
Daly, 1996[256]	Yes	Yes	Yes	Yes
Francis, 1995[250]	Yes	Yes	Yes	NA
Garton, 1997[257]	Yes	Yes	Yes	NA
Geelhoed, 1994[247]	Yes	Yes	Yes	Yes
OTA, 1995[243]	Yes	Yes	Yes	NA
Rosner, 1998[260]	Yes	Yes	Yes	Yes
Torgerson, 1993[251]	No	No	No	NA
Torgerson, 1995[252]	Yes	Yes	Yes	NA
Tosteson, 1990[248]	Yes	Yes	Yes	No
Tosteson, 1991[249]	Yes	Yes	No	No
Visentin, 1997[253]	Yes	No	No	NA
Weinstein, 1980[258]	Yes	Yes	Yes	Yes
Weinstein, 1983[259]	Yes	Yes	Yes	Yes
Weinstein, 1990[262]	Yes	Yes	Yes	Yes

NA, not applicable.

TABLE 25 *Relevance of the model input to the stated decision-maker*

Study (year)	Q3.2a Clinical data sources relevant to decision-maker?	Q3.2b Resource use data relevant to decision-maker?	Q3.2c Unit costs relevant to decision-maker?	Q3.2d Preferences/ utilities relevant to decision-maker?
Ankjaer-Jensen, 1996[245]	Yes	Yes	Yes	NA
Cheung, 1992[254]	No	Yes	Yes	Not clear
Coyle, 2001[261]	Yes	Yes	Yes	Yes
Daly, 1992[255]	Yes	Yes	Yes	Not clear
Daly, 1996[256]	Yes	Yes	Yes	Yes
Francis, 1995[250]	Not clear	Not clear	Not clear	NA
Garton, 1997[257]	Not clear	Not clear	Not clear	NA
Geelhoed, 1994[247]	Yes	Yes	Yes	Not clear
OTA, 1995[243]	Yes	Yes	Yes	NA
Rosner, 1998[260]	Yes	Yes	Yes	Not clear
Torgerson, 1993[251]	Not clear	Not clear	Not clear	NA
Torgerson, 1995[252]	Not clear.	Not clear	Yes	NA
Tosteson, 1990[248]	Not clear	Yes	Yes	Not clear
Tosteson, 1991[249]	Not clear	Yes	Yes	Not clear
Visentin, 1997[253]	In part	Not clear	Not clear	NA
Weinstein, 1980[258]	Yes	Not clear	Not clear	Not clear
Weinstein, 1983[259]	Yes	Yes	Yes	Not clear
Weinstein, 1990[262]	Yes	Yes	Yes	Not clear

NA, not applicable.

Resource use data and unit costs

The sources of resource use were explicit for the majority of the studies included in the review (*Table 24*). For example, resource use estimates and assumptions were explicit in the study reported by Ankjaer-Jensen and colleagues[245] and Rosner and colleagues used a Delphi panel to estimate resource use.[260] Some studies included drug costs only in the estimate of resource use.[250–253] Sixteen studies in the review reported most sources of unit costs incorporated in the analyses whereas the remaining two studies[251,253] did not report any sources for unit cost data.

Not surprisingly, it was difficult to judge whether the estimates of resource use and unit costs were relevant to the decision-maker for those studies that did not explicitly state the decision-making audience for the study (*Table 25*). For those studies for which a decision-maker was explicit or could otherwise be inferred, the sources of resource use were largely judged to be relevant. The target decision-maker was not clear in the studies by Tosteson and colleagues[248,249] and Weinstein;[258] however, the resource use would have been relevant provided that the target decision-maker was Medicare. It was not clear whether resource use was relevant to the decision-maker in four studies.[250,251,253,257]

Sensitivity analysis on context-specific factors

Clinical estimate and health state valuation

The studies were also reviewed to assess whether they had considered the implications of variation in input parameters using sensitivity analysis. Ten studies (55%) explored alternative assumptions of effect estimate in sensitivity analyses. Ankjaer-Jensen and colleagues[245] used "best case" and "worst case" scenarios in their analysis of effectiveness, and Rosner and colleagues[260] explored the 95% CI boundaries for vertebral fracture rates in the model. Similarly, Cheung and Wren[254] varied the risk of death from myocardial infarction (MI) over a range and the OTA[243] varied the risk of all clinical parameters (bone loss, cancer and heart attack) in its model. Daly and colleagues[255,256] varied both the magnitude and the duration of effect estimates in their sensitivity analysis, whereas Geelhoed and colleagues[247] explored the impact of HRT on different body systems (e.g. breast cancer) in the sensitivity analysis. The study by Coyle and colleagues[261] based its effect estimate on a meta-analysis of several trials and found that the cost-effectiveness estimate was highly sensitive to the inclusion of one particular study.

These sensitivity analyses were largely undertaken to explore **parameter uncertainty** (e.g. due to sampling uncertainty) rather than explicitly to consider possible **variability** in clinical effects within or between jurisdictions. In part, this comment also applies to the two models which estimated hip fracture rates from BMD.[248,262] However, in varying the population baseline hip fracture risk in these studies (for example, baseline hip fracture risk was increased by 100% and decreased by 50% in Tosteson and colleagues' study[248]), this would have been of interest to decision-makers as adjustment of baseline risk is often used to adapt the results of models between geographical areas (see Chapter 4).

Assumptions of compliance and duration of treatment were made without adjustment to clinical practice circumstances in the majority of studies in the review. For example, the OTA[243] assumed 100% compliance over 10, 20, 30 and 40 years; Daly and colleagues[256] assumed 100% compliance over 5, 10, 15 and 20 years. The definition of compliance differed between the eight studies which took this into consideration, but often it meant simply that patients "declined to accept" therapy[254] or that patients "accepted but discontinued" therapy[255] (*Table 26*). Generally, the cost-effectiveness estimates were found to be sensitive to the assumption of compliance, but the recommended policy-decision of the studies remained unchanged (*Table 26*). For example, Tosteson and colleagues[248] assumed 100% compliance over 15 years in the base case model, but varied compliance to 30% in the sensitivity analysis and found that cost-effectiveness estimates were sensitive to assumption of compliance.

One study[261] explored the sensitivity of health state valuation. Alternative utilities that assigned a 0.1 higher utility weight to the postfracture health state for women receiving nasal calcitonin were identified. The estimates of ICER increased as a result of the alternative utilities. Again, however, this sensitivity analysis was associated more with parameter uncertainty than geographical variation.

Only five studies contrasted their findings with other economic evaluation studies in the area (*Table 27*). For example, the report by the OTA[243] provided a comprehensive discussion of methodology, costs, clinical assumptions and results in relation to other cost-effectiveness analyses. In principle, this would have allowed decision-makers to assess whether other studies in the field had incorporated more appropriate data inputs for their jurisdiction.

Resource use and unit costs

Most studies applied unit cost data to the analysis relevant to the country for which the study was targeted (*Table 28*). Three studies used regional cost estimates.[247,254,260] None of the models accommodated differences in treatment patterns within regions or across countries. Furthermore, none of the investigators attempted to use cost estimates applicable to a broader audience of decision-makers within or between countries by using a range of costs or treatment patterns representing geographical differences. The

TABLE 26 *Definition of patient compliance and impact on compliance on the results of the cost-effectiveness analysis*[a]

Study	Definition of compliance	Result of analysis
Ankjaer-Jensen, 1996[245]	Compliance was % of patients attending BMD screening. Compliance of 100% and 50% was explored	Screening was more cost-effective than population-based approach under both assumptions of compliance (attendance)
Cheung, 1992[254]	Compliance was % of patients completing 15 years of treatment. Model base-case was 90% compliance for high-risk patients and 70% compliance for mid-risk patients. Explored impact of reduced compliance at 10% intervals	The estimate of cost savings was more sensitive to compliance in high-risk group. "Break even" (cost of savings = cost of intervention) was reached at compliance of 77% in high-risk only or 70% in high-risk and 40% in mid-risk group
Coyle, 2001[261]	Compliance was proportion of patients taking medication after 1 year. Base-case was 50% after 1 year. The impact of 25% and 75% compliance was tested in addition to 10% superior compliance with calcitonin	Analysis was sensitive to the assumption about compliance with therapy. However, the overall results of the analysis did not change
Daly, 1992[255]	Compliance was % of patients continuously taking the drug. Compliance of 100% in the first 5 years falling to 50% (oestrogen only, ORT) and 67% falling to 33% (oestrogen and progestin)	Cost-effectiveness increased with reduced compliance over all treatment strategies. The overall results of the analysis did not change
Garton, 1997[257]	Compliance rate used implied % patients initiating therapy and continuing beyond year 1 to complete a course of 10 years of treatment. Compliances of 10, 30 and 50% were explored	Universal HRT was more cost-effective than screening strategy both under high and low compliance. However, if screening could increase compliance then screening could prove more cost-effective
Geelhoed, 1994[247]	Assumed that patients who fill prescriptions but do not take the drug incur costs but gain no benefits. Explored scenario where 70% of prescriptions are filled but only 30% are taken as directed	The net cost per QALY would increase ("more than double") under a scenario of reduced compliance
Rosner, 1998[260]	Different rates of willingness of patients to initiate (WTI) and continue (WTC) treatment were incorporated in the model ranging from 18.1 to 100%. These were based on epidemiological studies	The model was "moderately sensitive" to changes in WTI and WTC. One strategy was particularly sensitive but it remained cost-effective under the assumption that public willingness to pay for a QALY is >Can$100,000
Torgerson, 1993[251]	Compliance rates were defined as 'willingness to initiate' therapy. Once therapy initiated, 100% compliance assumed in 30 years. Different rates of willingness of patients to initiate and continue treatment were varied from 18.1 to 100% and based on epidemiological studies	The outcome of the evaluation was sensitive to the assumption of compliance. More than 50% need to initiate therapy ("be compliant") in order that targeted intervention is cost-effective
Tosteson, 1990[248]	Assumed 100% compliance with treatment varied from 5 years to lifetime use (baseline model 15 years). Explored 30% compliance in sensitivity analysis	Results were "sensitive", but compliance did not change the main conclusion of the model: screening remained more cost-effective than universal treatment

[a] Only studies explicitly considering compliance are included in the table.

sensitivity of the study results to national variation in cost per fracture was only explored in the study reported by Coyle and colleagues,[261] who found that their cost-effectiveness estimate was markedly reduced when using cost calculations from an alternative cost-of-illness study.

Authors' comments on generalisability

Overall, four studies explicitly commented on the issue of the generalisability of their analysis to other settings in their presentation or discussion of results (*Table 29*).[245,251,259,261] For example, Coyle and colleagues[261] commented

TABLE 27 *Sensitivity analysis of estimates of clinical effectiveness and health state valuation*

Study (year)	Q4.1a Was the robustness of the effect estimate explored?	Q4.1b Did the model accommodate for potential difference in compliance rates?	Q4.1c Did the model accommodate for differences in utilities?	Q4.1d External consistency: were the results compared with other relevant studies?
Ankjaer-Jensen, 1996[245]	Yes	Yes	NA	Yes
Cheung, 1992[254]	Yes	No	No	No
Coyle, 2001[261]	Yes	Yes	Yes	No
Daly, 1992[255]	Yes	Yes	No	Yes
Daly, 1996[256]	Yes	No	No	No
Francis, 1995[250]	No	No	NA	No
Garton, 1997[257]	No	Yes	NA	Yes
Geelhoed, 1994[247]	Yes	Yes	No	No
OTA, 1995[243]	Yes	No	NA	Yes
Rosner, 1998[260]	Yes	Yes	No	No
Torgerson, 1993[251]	No	Yes	NA	Yes
Torgerson, 1995[252]	No	No	NA	No
Tosteson, 1990[248]	Yes	Yes	No	No
Tosteson, 1991[249]	No	No	No	No
Visentin, 1997[253]	No	No	NA	Yes
Weinstein, 1980[258]	No	No	No	NA
Weinstein, 1983[259]	No	No	No	No
Weinstein, 1990[262]	Yes	No	No	No

TABLE 28 *Sensitivity analysis of resource use and valuation*

Study (year)	Q4.2a Did the model reflect variation in costs nationally?	Q4.2b Did the model reflect variation in resource use patterns nationally?	Q4.2c Did the model reflect variation in costs internationally?	Q4.2d Did the model reflect variation in resource use patterns internationally?
Ankjaer-Jensen, 1996[245]	No	No	No	No
Cheung, 1992[254]	No	No	No	No
Coyle, 2001[261]	Yes	No	No	No
Daly, 1992[255]	No	No	No	No
Daly, 1996[256]	No	No	No	No
Francis, 1995[250]	No	No	No	No
Garton, 1997[257]	No	No	No	No
Geelhoed, 1994[247]	No	No	No	No
OTA, 1995[243]	No	No	No	No
Rosner, 1998[260]	No	No	No	No
Torgerson, 1993[251]	No	No	No	No
Torgerson, 1995[252]	No	No	No	No
Tosteson, 1990[248]	No	No	No	No
Tosteson, 1991[249]	No	No	No	No
Visentin, 1997[253]	No	No	No	No
Weinstein, 1980[258]	No	No	No	No
Weinstein, 1983[259]	No	No	No	No
Weinstein, 1990[262]	No	No	No	No

that the results were sensitive to the baseline population hip fracture risk, and for that reason the generalisability of the results were unclear.

Discussion

This chapter has reviewed the use of decision analytic cost-effectiveness models, in a specific

TABLE 29 *Authors' comments on the generalisability of the results*

Study (year)	Did the author explicitly address the issue of transferability of the results to other jurisdictions?
Ankjaer-Jensen, 1996[245]	Yes. The authors commented that the results are limited to a Danish context, considering wide variability in drug costs across countries. For example, they comment that drugs cost 50% less in Sweden and that this would influence the results
Cheung, 1992[254]	No
Coyle, 2001[261]	Yes. The authors commented that the results are sensitive to baseline risks of fracture in the population and that the generalisability to other jurisdictions therefore was unclear
Daly, 1992[255]	No
Daly, 1996[256]	No
Francis, 1995[250]	No
Garton, 1997[257]	No
Geelhoed, 1994[247]	No
OTA, 1995[243]	No
Rosner, 1998[260]	No
Torgerson, 1993[251]	No
Torgerson, 1995[252]	Yes. The authors commented that the cost of hip fracture was excluded because the hospital length of stay varies between UK regions and therefore an estimate based on a region with long stay would not apply to an estimate in a region with short stay
Tosteson, 1990[248]	No. However, the authors acknowledged that baseline fracture risk in the population is important to the cost-effectiveness ratio
Tosteson, 1991[249]	No
Visentin, 1997[253]	No. However, the authors noted that the high cost-effectiveness ratio may have been due to relatively low incidence of hip fracture in the Italian population compared with other populations
Weinstein, 1980[258]	No
Weinstein, 1983[259]	Yes. The authors stated that the cost advantage of the estrogen–progestin alternative would have been even greater if British recommendations on more frequent endometrial monitoring had been adopted
Weinstein, 1990[262]	No

clinical area, to assess a range of factors associated with the potential relevance of an analysis to a particular decision-maker and the extent to which its results might be transferable to other jurisdictions. Although, as part of this process, there has been a consideration of how models are presented, this is not a general critical review of modelling methods. More general issues of good methods in decision analytic cost-effectiveness modelling have been considered elsewhere.[280,286] A general critical review of cost-effectiveness models in osteoporosis has also been published.[241]

The purpose of cost-effectiveness decision models is to aid decision-makers to make appropriate resource allocation decisions given available data. As such, it would seem reasonable for any analysis to be clear about the jurisdiction and specific decision maker for which the model is designed. However, only 3 (16%) of the studies in this review provided these details explicitly, although it was possible to infer the decision context from other information provided. Similarly, authors frequently report the methods and results of studies for which a firm research question has not been stated. To aid decision-making, models also need to be clear about the decision problem(s) being addressed. The majority of reports in the review defined the study setting, patient population, model structure and key assumptions in a transparent manner.

Once the target decision-maker/jurisdiction and decision problem have been established, the former will need to decide whether the data inputs and assumptions in the model are the best available for their context. The majority of the reports provided sources for clinical and economic data and their valuation. The data inputs to those models for which a decision-maker was stated or could be inferred appeared to be relevant and, as far as could be judged from only the study report, were the 'best available' for the decision context. As mentioned in earlier chapters, there is often an implicit assumption in models that estimates of clinical effectiveness are transferable between locations in a way that resource use and cost data are not. Perhaps reflecting this assumption, the papers in the review generally made more effort to ensure (and to be seen to ensure) that their cost inputs (at least unit costs) were specific to their

target jurisdiction. Most papers were prepared to use clinical data from studies undertaken outside their context. As in other clinical areas (see Chapter 4), an exception to this assumption of the transferability of clinical data is the adjustment of baseline risks to make them specific to a particular jurisdiction (usually country) whilst assuming that the **relative** treatment effect is exchangeable geographically. This adjustment of baseline risks was undertaken rarely in the sample of papers reviewed,[248] although the effect of variation in these parameters for the generalisability of analyses was discussed in another paper.[261]

Little attempt was made to justify the particular utilities used in those cost–utility models in the sample with respect to the target decision-maker. In part, this probably reflects the limited amount of utility data available relating to osteoporosis. In other words, the authors often used any utility data which were available – in the majority of studies this was the assumptions (rather than empirical elicitation) made by Weinstein.[258] Increasingly, decision-makers will be specific about the type of health state utility data they wish to see in economic evaluations. For example, NICE has indicated that it wishes to see the use of public preferences relating to the English and Welsh population in cost–utility analyses submitted to its technology assessment programme.[6] There are two aspects to this decision maker preference. The first is the position a given decision-maker takes on the most appropriate utility estimation methods (e.g. use of patient or public preferences). The second is the issue of whether the preferences of individuals outside the particular jurisdiction of interest are acceptable.

One aspect of the review was to assess the extent to which studies had assessed the impact of variability in parameter estimates associated with location using sensitivity analysis. In principle, this sort of analysis might be undertaken for two reasons. Firstly, there may be variability in clinical and/or cost parameters **within** a given jurisdiction. If there was good reason to think that this level of variability might impact on the conclusions of the analysis, sensitivity analysis would be strongly indicated. The second way in which sensitivity analysis could be used is to assess whether the results of the model, as they apply to the target jurisdiction, would also apply to other locations by appropriate variation in parameter values. If it is accepted that the purpose of decision models is to address decision problems for particular jurisdiction/decision-makers, then this process of generalisation should probably not be seen as an essential element of model-based economic analysis.

Although most of the studies in the review undertake extensive sensitivity analysis, few explicitly do this to explore variation between locations/jurisdictions. Only one[243] of the models is probabilistic, that is, reflecting the uncertainty in parameters as random distributions and propagating that uncertainty through the model, to be jointly reflected in the results, using Monte Carlo simulation.[11] Therefore, most of the standard one-way or two-way sensitivity analysis was undertaken to assess the importance of parameter uncertainty to the results, rather than variability between locations. The failure to assess variability within and between locations may have reflected the view that parameter uncertainty dwarfed variability within the target jurisdiction, and generalising their results across locations was not a primary concern.

One area of interest in the review was how the models dealt with the issue of patients' compliance with therapy. This is important in terms of generalisability in that compliance is expected to vary between research (e.g. trial) settings and routine practice (see Chapter 3).[44] Different definitions were used by the authors in this review, but for the most part, the assumption of compliance was based on the fraction of patients which initiates therapy. The overall results of the analysis were sometimes substantially influenced by the assumption of patient compliance.

Conclusions

This chapter has looked at how issues of variability in cost-effectiveness across locations is handled in modelling studies in the field of osteoporosis. The review found that most studies either stated their target decision-maker/jurisdiction or provided sufficient information from which this could be inferred. There was a greater tendency to ensure that cost (resource use and unit cost) inputs were specific to the target jurisdiction than clinical parameters. Although there was extensive sensitivity analysis undertaken in most studies to assess parameter uncertainty, there was little use of these methods to explore the implications of variability within or between locations in parameter inputs. Only four studies explicitly commented on the issue of the generalisability of their analysis to other settings in their presentation or discussion of results.

Chapter 8

Case study II: making economic evaluation results specific to a target jurisdiction using decision modelling. The case of glycoprotein IIb/IIIa antagonists for acute coronary syndrome

As discussed in earlier chapters, the purpose of economic studies based on decision analytic models is to generate estimates of cost-effectiveness, based on available (uncertain) data, to inform resource allocation decisions. Given their focus on particular decision problems, decision models would be expected to be specific to a target decision-maker or jurisdiction. These would typically be at a national (or health care system) level (e.g. NICE in the UK), but they could also focus on a more specific decision-making body such as a hospital formulary committee. These models may then seek to generalise from the particular target decision-maker, but this is not their primary purpose, and Chapter 7 showed that this was rarely undertaken, at least in the field of osteoporosis.

The review in Chapter 7 also assessed the extent to which applied modelling studies in osteoporosis incorporated data inputs which were appropriate for the target jurisdiction or decision-maker as stated or inferred from each study. It was found that studies tended to be more assiduous in selecting cost inputs which were specific to their target decision-maker than they were in identifying appropriate clinical inputs. This is likely to reflect an implicit assumption that parameters relating to clinical effectiveness, although needing to be specific to the relevant patient group defined in the decision problem, are inherently more transportable geographically. Although this assumption may be justified **within** healthcare systems and countries, the factors reviewed in Chapter 3 suggest that this will not necessarily be so **between** systems and countries.

Chapter 3 shows that conceptually – and to some extent empirically – it would be expected that both the effectiveness and the cost side of economic evaluations will be affected by location. Therefore, model-based economic studies need to be able to show that best available clinical and cost inputs have been used **for the target decision-maker or jurisdiction**. In many instances the data which are available to model decision problems will come partly or wholly from sources outside the target decision-maker's location. This raises questions about the most appropriate methods for adapting the model and its data inputs to be most relevant for the particular decision problem and decision-maker. The aim of this chapter is to address some of these issues using the case study of a recent model developed for resource allocation decisions by NICE relating to the NHS.[263] In particular, the chapter describes approaches that can be adopted to adjust rates of baseline clinical events and to assess whether to adjust relative treatment effect. It also illustrates the use of sensitivity analysis to assess the implications of possible changes in clinical practice in the target jurisdiction, hence addressing the issue of variability in cost-effectiveness results over time.

Methods

Summary of the case study and decision problem

The case study relates to a decision model developed to assess the cost-effectiveness of glycoprotein IIb/IIIa antagonists (GPAs) in the management of non-ST-elevation acute coronary syndromes (ACSs). It is not the purpose of the chapter to provide full details of the clinical context or of the decision model, which can be found elsewhere.[263] The following provides a brief summary with the aim of providing a sufficient background to the consideration of the main issues relating to the use of models to apply international data to a specific jurisdiction.

Non-ST-elevation ACS includes either unstable angina or non-Q-wave MI. Non-Q-wave MI is the term used when the cardiac enzymes are elevated to the range indicating that MI has occurred, but a Q-wave does not develop on ECG tracings. Unstable angina itself represents a spectrum of

clinical states that falls between stable angina and acute MI. It includes new onset angina and angina occurring >24 hours post-MI.

GPAs are a new class of drugs that may be more effective in preventing the platelet aggregation associated with ACS than existing therapies such as aspirin and heparin.[264] Two broad groups of GPAs are licensed in the UK: abciximab (ReoPro®, Eli Lilly) is a monoclonal antibody targeted at the receptor (also known as a 'large molecule' GPA), whereas eptifibatide (Intergrilin®, Schering Plough) and tirofiban (Aggrastat®, MSD) are more conventional pharmacological receptor antagonists (also known as 'small molecule' GPAs).

The potential cost-effectiveness of GPAs relies on the extent to which, by reducing rates of mortality and MI, they are able to generate enough gain in quality-adjusted survival for the average patient and/or to reduce 'downstream' healthcare costs sufficiently to justify their acquisition cost. GPAs are used in two general ways to manage ACS patients. First, as an adjunct to percutaneous coronary interventions (PCIs) (e.g. coronary angioplasty) for those patients who undergo such a procedure – abciximab is the GPA which is mainly used for this purpose. Second, GPAs can be used as a form of medical management for non-ST-elevation ACS patients regardless of whether or not they subsequently undergo a PCI – tirofiban and eptifibatide are mainly used for this indication.

The evidence base relating to the **clinical efficacy** of GPAs is extensive, albeit partial.[264] Trials of medical management have randomised over 30,000 patients, and typically compare GPAs with standard management. Overall, these trials have shown a reduction in the risk of non-fatal MI or death – an odds ratio of 0.91 (95% CI 0.84 to 0.98) in a recent patient level meta-analysis.[265] Trials of the use of GPAs alongside PCI have been more heterogeneous in their intake, with only one trial focusing solely on non-ST-elevation ACS patients. Overall, studies recruiting some patients with ACS (10 trials randomising over 15,000 patients) have shown the use of GPAs to generate a relative risk of non-fatal MI of 0.68 (95% CI 0.57 to 0.80) and of death of 0.80 (95% CI 0.60 to 1.09) by between 30 days and 6 months.[264] Some economic evidence exists on GPAs,[264] part of which is based on the analysis of patient-level data alongside specific RCTs. However, none of these studies has adopted a long-term time horizon, expressed outcomes in terms of generic measures of health gain such as QALYs or compared a full range of feasible strategies for the use of GPAs in ACS. Most important for the purposes of this case study, none has focused on UK costs and clinical practice. As such, these studies provide minimal assistance for UK decision makers concerned with the reimbursement of GPAs.

Hence a decision model was developed to assess the cost-effectiveness of the four strategies detailed above in the context of NHS practice. Adopting a 50-year time horizon and health outcomes in terms of QALYs, the model was made up of two parts: a short-term element, which related to a period of 6 months after a patient presents with non-ST-elevation ACS, and a long-term element, which estimated a patient's lifetime costs and outcomes conditional on surviving the first 6 months after the acute episode. The model was probabilistic,[11] and Monte Carlo simulation was used to propagate second-order uncertainty in input parameters through the model to be reflected in decision uncertainty. A 2000–01 price base was used and annual discount rates of 6% for costs and 2% for benefits were adopted based on UK guidance.[266]

Relating the model to UK practice

The reason for selecting the GPA model as a case study is that, in building the model, a number of issues relating to the transferability of data across international boundaries had to be addressed. In particular, the principle source of data for clinical effectiveness and resource use was a set of randomised trials which were undertaken largely or wholly outside the UK. In identifying data inputs to incorporate into the model, it was necessary to balance the strengths and weaknesses of the available trial data. It was important for the selection of data in the model to exploit the internal strengths of the trial data stemming from the studies' experimental design and large sample sizes (about 50,000 patients in total in relevant trials). However, it was also essential to recognise the fact that data on resource use and clinical effects gathered in the trial may not reflect routine practice in the NHS. The following sections describe the limitations of the trial data and how these were addressed in the decision model. *Table 30* summarises the limitations of the data from the trials for a UK-focused analysis and the methods used to overcome these.

Does the use of GPAs in trials reflect their use in routine NHS practice?

In developing the UK model, an important consideration was how GPAs would be used in routine clinical practice. As described above, the

TABLE 30 *Summary of the limitations of the GPA trials for a cost-effectiveness model for the UK and approaches taken to overcome these*

Limitation with trial data	Methods used to overcome limitations	Additional data source used
Use of GPAs in the trials does not reflect their range of possible uses in the UK	Four strategies relating to how GPAs are being used in the UK were identified with clinical collaborators	
Baseline event rates in trials are unlikely to reflect UK practice owing to differences in management of ischaemic heart disease	Separate out baseline event rates in trial control groups from relative treatment effects Apply latter to UK-specific baseline event rates taken from other sources	UK-specific baseline event rates taken from (PRAIS-UK)[269] and a specific survey of patients undergoing acute PCI at Leeds General Infirmary
The relative treatment effect may also vary by location. In particular, the relative risk may be related to the baseline risk in a location	Undertake a meta-regression to relate the baseline (control group) risks with those in the experimental (GPA) groups. If a clear relationship is identified, this can be used to adjust the relative risks used for a UK analysis to a level commensurate with UK baseline event rates	
Recent changes in clinical practice in the UK include the increased use of PCI and use of clopidogrel as part of standard management. However, the UK data sources may not reflect these changes, and their implications for the cost-effectiveness of GPAs are unclear	Use of sensitivity analysis to assess the effect on the models' results of using baseline event rates from the trials instead of the UK observational study. Also modelling a strategy of using clopidogrel as a fifth strategy in the model	
No data from the trials (UK-specific or otherwise) to extrapolate short-term clinical results to long-term QALYs	Use of a long-term Markov model populated using UK-specific observational data	Transition probability and longer term resource use data taken from NHAR[275]

evidence base contains two types of GPA trial: those comparing the drugs with standard practice (i.e. management without GPAs) in all patients with non-ST-elevation ACS regardless of whether PCI was subsequently undertaken (medical management); and those looking at GPAs as an adjunct to PCI. However, it was possible to identify four strategies relating to how these drugs were being used in clinical practice in the UK, as outlined below:

- Strategy 1: GPA as part of initial medical management. This involves patients with ACS receiving an infusion of GPA as soon as their 'high-risk' nature has been established.
- Strategy 2: GPA in patients with planned PCIs. GPA is started once a decision to undertake PCI (or angiography with a view to proceeding to PCI) has been made.
- Strategy 3: GPA as adjunct to PCI. GPA is used at the time of PCI or is started up to 1 hour before the procedure in those patients undergoing such a procedure.
- Strategy 4: no use of GPA. With this strategy, patients are assumed to receive standard therapies (e.g. heparin, aspirin, nitrates and analgesia), without the use of GPA.

The model was, therefore, structured directly to compare these four strategies. Given that none of the trials directly compared all of these strategies, however, it was necessary to restructure the effectiveness data to reflect the nature of the indirect clinical comparison which was needed to populate the decision model. This was achieved by separating out the baseline event rates measured in the standard therapy control groups in the trials from the treatment effect observed in the GPA arms relative to the control group. The relative treatment effects were pooled across the various groups of trials. The relative treatment effect for Strategy 1 was based on seven medical management trials randomising a total of 30,280 patients; the relative risk reduction for Strategy 2 was taken from a single trial of 1265 patients; and the treatment effect of Strategy 3 was based on 10 trials randomising a total of 15,951 patients. In principle, the baseline event rates of interest (i.e. rates of death, MI and revascularisation) could have been taken from pooled control group data from trials. This could have been taken as representing event rates in Strategy 4 (no use of GPAs). For reasons described below, however, baseline event rates were actually taken from another source.

Are baseline event rates in the trials relevant to UK practice?

Given that the trials were undertaken largely outside the UK, the baseline event rates in patients not having GPAs in the UK may be different to those patients randomised to the control groups in the trials. This could reflect differences in the epidemiology of the disease or, more probably, differences in overall management of patients with ischaemic heart disease in the UK. Traditionally, the principal difference in the management of ischaemic heart disease in the UK, compared with that in other developed countries, is that fewer patients are considered for PCI.[267] This is important for two reasons. First, early use of PCI in ACS patients has been shown to reduce rates of death and MI in recent randomised trials,[268] hence lower rates of PCI in the UK may have the effect of generating higher baseline event rates than seen in the GPA trials. Second, the limited availability of 'acute' PCI (i.e. percutaneous procedures undertaken in non-ST-elevation patients shortly after presentation) in the NHS may cause clinicians to select ACS patients for acute PCI in a different way than clinicians in the GPA trials.

Therefore, in developing the decision model, baseline event rate data which were specific to UK practice were sought. These data were taken from the Prospective Registry of Acute Ischaemic Syndromes in the UK (PRAIS-UK).[269] This is an observational cohort registry of 1046 patients admitted to 56 UK hospitals with ACS in 1999. Patients were followed up for 6 months after their index hospital admission and the hospitals included in PRAIS-UK served 24% of the UK population. For the purposes of the study, patients who received GPA in PRAIS-UK (n = 13; 1%) were excluded from the analysis. The parameter estimates from PRAIS-UK relating to patients who received a PCI during the acute phase of their ACS were based on a relatively small number of patients (n = 53). For this reason, an audit of unstable angina patients undergoing acute PCI at a large UK cardiac centre (Leeds) was undertaken to supplement data from PRAIS-UK (n = 231).

Are the relative risk reductions estimated from the trials related to baseline risks?

One way of adapting the clinical results from international trials to the UK setting is by separating out the baseline event rates associated with standard management (without GPAs), estimating those parameters from UK-specific data and applying the pooled relative treatment effects, for Strategies 1–3 relative to Strategy 4, from the trials. This approach effectively assumes that baseline risks are not transferable internationally, but relative risk reductions are. It may be, however, that the relative treatment effect is itself related to baseline risk – for example, the higher the baseline risk, the lower is the treatment effect – in which case the assumed independence between the two components of clinical effectiveness would not be sustainable.

In order to investigate whether the log relative risk in the individual trials varied with log baseline risk (i.e. the log event rate in the control group), a random effects meta-regression model was used.[270] Initially this was performed using the metareg macro in STATA but, in order to adjust for regression to the mean associated with the dependent variable being a function of the independent variable, the method advocated by Sharp and Thompson was used,[271] implemented in WinBUGS.[272] In other contexts, this form of analysis has been used to look at the generalisability of the absolute treatment effect identified in trials across a range of clinically defined patient subgroups where the same separation of baseline risks and relative treatment effect can be employed.[19]

In effect, this form of meta-regression works by fitting a regression line between the event rates measured in the control groups of the trials and those of the experimental (i.e. GPA) groups, with the number of points from which the estimate is made being the number of available trials. This function characterises the relationship between the baseline risks and the relative risks, and this can be either positive or negative. This can then be used in the decision model to adjust the relative risk estimates according to the baseline risk employed in the model. In the context of the GPA trial, the value of this approach relates to the potentially different baseline risk in patients presenting with non-ST elevation ACS in the UK compared with those randomised into the international trials. Once the function has been estimated in the meta-regression, the pooled relative risk estimates from the trials could be adjusted according to the point on the regression line which accords with the UK baseline risk.

How can recent and possible future changes in UK practice be reflected in the model?

In developing cost-effectiveness models to inform UK decision-making, an important consideration is the need to try to reflect the speed with which clinical practice changes. Given the time necessary to analyse and publish trials and observational

studies, the data sources which are available for models are likely to reflect clinical practice which is at least 2 years old and may no longer accurately reflect contemporary clinical practice. This is a particularly important issue in the management of ischaemic heart disease in the UK. Although practice in the UK has been characterised by a low rate of revascularisation, routine data suggest this is changing. In particular, PCI rates in the UK have sharply increased in recent years.[267] Although the PRAIS-UK dataset which was used to generate UK-specific baseline data for the model was published as recently as 2000, the relevant data were taken from 1999, and these may be somewhat out of date.

To assess the likely importance of these developments to the cost-effectiveness of the alternative GPA strategies, therefore, sensitivity analysis was undertaken which explored how the results of the model varied when UK-specific baseline risks were exchanged for the baseline risks from the clinical trials which were taken from a patient-level meta-analysis.[265] The implicit assumption with this sensitivity analysis was that the baseline event rates in the clinical trials represent the sort of clinical practice towards which the UK is moving, so it was important to consider whether it was possible that the optimal decision suggested by the model would change as practice altered.

Another aspect of clinical practice that is likely to change in the near future is what constitutes 'standard practice' against which the strategies using GPAs should be compared. In the trials, patients in the control groups typically received heparin and aspirin as 'standard'. One possibility is that another medication – clopidogrel – will become widely used in the UK. Clopidogrel is an alternative antiplatelet agent, with a different mechanism of action to the GPAs. The CURE trial[273] indicated that clopidogrel was effective in combination with aspirin and guidelines are being modified accordingly.[274] Clopidogrel is also cheaper than the GPAs, although it is taken over a longer period. If clopidogrel becomes widely used as 'standard practice' in the UK, this may affect the cost-effectiveness of the GPAs. Therefore, a sensitivity analysis was undertaken to assess these implications using relative risk results from the CURE[273] trial to model a fifth strategy based in clopidogrel plus standard management.

What other UK-specific data need to be incorporated into the model?

An important function of many decision models is to extrapolate from effectiveness data available in trials. In the context of the GPA model, this extrapolation had two elements. The first was to extrapolate from the short-term effectiveness data in the trials, which typically had follow-up of no more than 6 months and frequently as short as 30 days, to the lifetime time horizon which is more appropriate for a potentially life-saving intervention used in patients with a chronic disease. The second form of extrapolation was in order to relate short-term data in terms of clinical effects (deaths, non-fatal MIs) to generic measures of health gain, necessary for resource allocation decision making, such as QALYs.

Therefore, a long-term (extrapolation) model was developed to estimate the future prognosis for patients who finish the short-term (6-month) model in one of two disease states: those having experienced a non-fatal MI and those who have not but remain alive. That prognosis will include the possibility of patients experiencing further non-fatal MIs in addition to dying for any reason. Hence the extent to which the use of GPAs reduces the risk of death and non-fatal MI, relative to baseline, during the initial 6-month period was translated into differences in long-term costs and QALYs on the basis of the long-term model.

The long-term model took the form of a four-state Markov process with states of ischaemic heart disease (that is, patients who had not experienced a non-fatal MI), non-fatal MI (where patients spend a single cycle of 1 year), post-MI (which surviving patients entered after 1 year following an MI) and death. In populating this model, it was necessary to be aware of the variation between countries in long-term survival following cardiac events. Transition probabilities were, therefore, taken from a UK-specific observational study – the Nottingham Heart Attack Register (NHAR).[275] Two cohorts of patients (total $n = 1279$) from the NHAR were used, with a diagnoses indicative of ACS, which had follow-up data for up to 5 years. It was not possible, however, to identify UK-specific quality-adjustment data in the literature. US data were used for this purpose,[276] with a constant decrement applied to all living patients in the model. Extensive sensitivity analysis was necessary to assess the extent to which the value of this 'QoL decrement' was crucial to the optimal management strategy identified by the model.

As indicated in the review of applied decision models in Chapter 7, it would be expected in most models that estimated costs would be based wholly or partly on data specific to the target decision-

maker or jurisdiction. With decision models, the process of estimating overall costs typically involves attaching cost weights to particular events, where the latter may actually be based on clinical data from outside the jurisdiction. Costing events requires a combination of information on resource use and unit costs (prices). For example, a model may generate estimates of the rate of MI, to which costs are then attached. In costing an MI, it would be necessary to estimate resource use such as days in hospital, use of medications and selection of diagnostic tests; then each of these items would need to be appropriately costed. In the context of the GPA model, the relevant 'events' estimated in the short-term model included revascularisation rates and days in hospital. For the longer term model, these events included MIs, revascularisations and days in hospital. In both cases, these estimates of event rates were based on data from the UK observational studies – PRAIS-UK for the short-term model and NHAR for the long-term model. All estimates of the costs of these events were then taken from UK sources.[263]

Results

Base-case results

Full details of the results of the GPA model and all sensitivity analyses can be found elsewhere.[263] Here we provide details of the base-case results of the study and of the sensitivity analyses which relate specifically to the methods which were used to make the model inputs specific to the UK.

The base-case analysis implemented all the methods discussed in the previous section to reflect UK data as fully as possible. This included the use of UK-specific baseline risks, costs and data relating to long-term extrapolation. The meta regression examining the relationship between the relative risks and baseline risks in the trials was undertaken, and an example is shown in *Figure 5* based on the estimate of mortality rates at 30 days in the four Strategy 1 trials for which data were available. This suggests a negative relationship between the log baseline and log relative risks – that is, the higher the log baseline (control group)

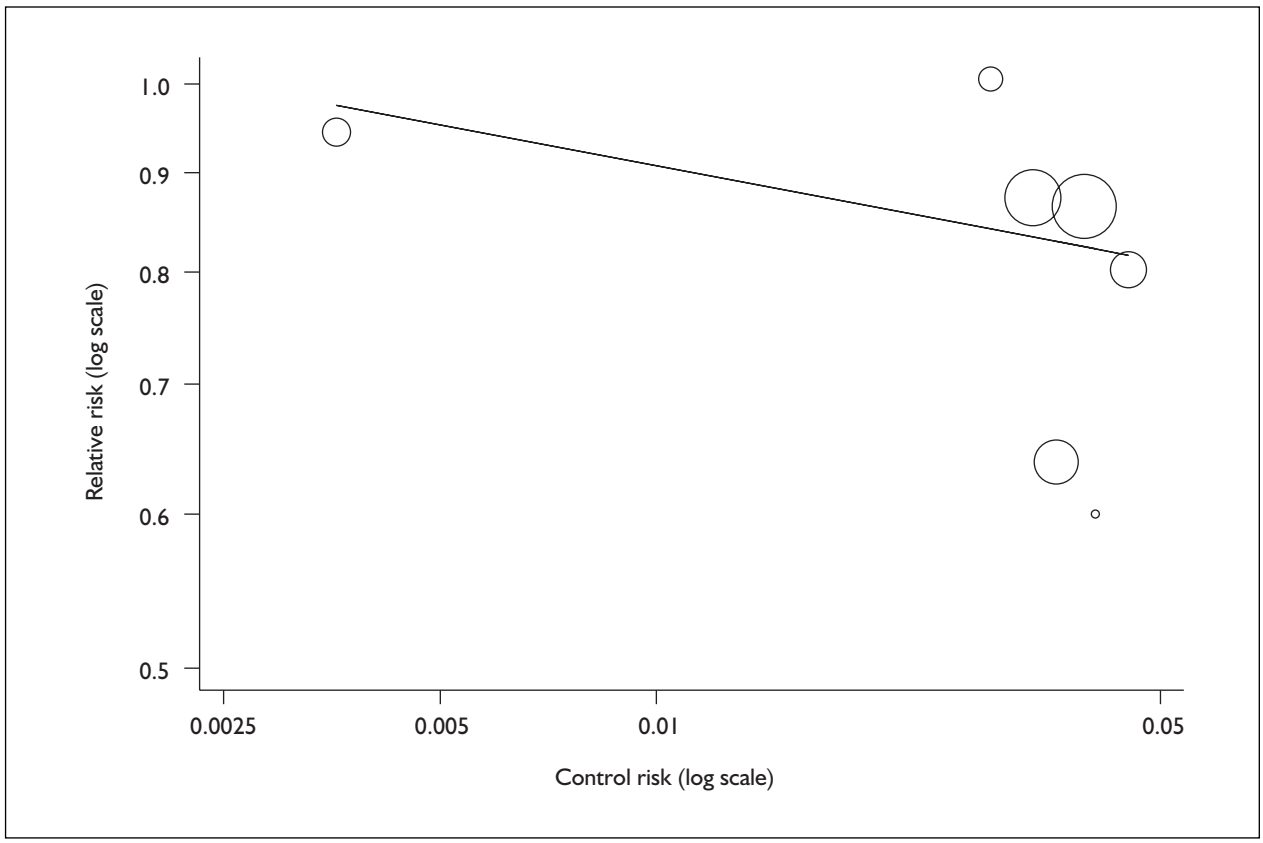

FIGURE 5 *Example of the results of the meta-regression comparing log baseline (control group) risks with log relative risks. This example relates to Strategy 1 trials and the event of mortality at 30 days. Each circle represents a trial and the size of the circles indicates their relative sample size. In this example, a constant (standard error) of –0.418 (0.524) and a slope (standard error) of –0.066 (0.150) were estimates; this showed that the slope was not statistically different from zero (p = 0.44).*

TABLE 31 *Base-case estimates of mean lifetime costs and QALYs for the four strategies, together with incremental analysis*

				Probability cost-effective for maximum WTP[a]		
Strategy	**Cost (£)**	**QALY**	**ICER**	**£10,000**	**£30,000**	**£50,000**
1	12,688	7.7875	£5738	81.67	94.15	95.19
2	12,207	7.6839	D	0.48	0.6	0.53
3	12,188	7.6910	ED (£25,811)[b]	1.03	2.77	3.01
4	12,119	7.6883		16.82	2.48	1.27

[a] The probability that each strategy is more cost-effective than the others conditional on different maximum willingness to pay for an additional QALY.
[b] ICER Strategy 3 versus Strategy 4.
D, dominated; ED, option ruled out by extended dominance.

risk, the lower are the log relative risks. However, the results of these meta-analyses were not implemented in the model for two reasons. First, none of the estimated relationships reached statistical significance. Second, and more importantly, the small number of trials (seven for 30 days event rates for Strategy 1) made the estimated relationship potentially unreliable.

The base-case results shown in *Table 31* indicate that, under base-case assumptions, Strategy 1 – the use of GPAs as a medical management early in presentation regardless of whether the patient subsequently went on to have a PCI – was the most cost-effective strategy assuming that the UK health service is willing to pay at least £5667 per additional QALY. Strategy 2 was dominated (lower effects and higher costs than other strategies) and Strategy 3 is subject to extended dominance[277] – that is, a higher incremental cost-effectiveness ratio than a more effective strategy. The uncertainty in this decision is presented in the table in terms of the probability that a strategy will be the most cost-effective given the decision-maker's willingness to pay for a QALY gained in these patients.[206] At a willingness to pay £30,000, for example, Strategy 1 has a 94% probability of being the most cost-effective.

Sensitivity analysis I: changes in baseline data

New baseline event probabilities were derived from the control group data of Boersma and colleagues' patient level meta-analysis of Strategy 1 trials.[265] Since the meta-analysis only reported event rates at 30 days, an extrapolation was required in order to apply these data to our short-term decision model since the model requires event rates at 6 months. This used the predicted hazard of these events estimated from the Strategy 1 trials reporting at both time intervals.

In the meta-analysis, patients were categorised into those who underwent acute PCI, acute coronary artery bypass graft (CABG) and no acute intervention depending on whether patients had undergone an intervention within 5 days of randomisation. Using these data, the baseline rate of PCI increases from 5%, in the base-case model based on UK observational data, to approximately 15% using the data from the trial control groups. The rate of acute CABG increased only marginally from 4.5% to 4.9%.

Table 32 summarises the baseline event rates for death and non-fatal MI using the alternative data sources for the three patient groups considered in the short-term model. Both the 30-day event data reported by Boersma and colleagues and the extrapolated event data at 6 months have been provided to illustrate the impact of the assumptions used in the extrapolation on each of the relevant events. The effect of changing the source of baseline event data appears to have the largest impact on the event rates reported in the acute PCI group: the rate of death increases from 3.3% using the UK specific baseline data to 5.62% using the meta-analysis. Similarly, the rate of non-fatal MI rises from 3.6% to 19.27%. In both data sources the death rate in the acute PCI group at 6 months is lower than that in patients who do not undergo acute revascularisation (although this differential is reduced using the new baseline data). However, the rate of non-fatal MI is now higher in the acute PCI group using the new baseline data (19.27% versus 13.43% compared with 3.6% versus 4.7%).

Two separate sensitivity analysis have been undertaken using the new baseline event data. The first analysis applies the same relative risks as in the base case based on a non-patient-level meta-analysis undertaken on all trials (i.e. relating to Strategies 1, 2 and 3).[263] The second analysis

TABLE 32 *Comparison of baseline event rates between UK-specific sources used in the base-case model and control group data from the trials derived from Boersma and colleagues*[265]

Revascularisation group	Event	Source of baseline event for the trial		
		Base-case model – 6-month event rates from UK-specific sources (%)	Boersma meta-analysis of trials – 30-day event rates (%)	Boersma meta-analysis of trials – 6-month event rates (extrapolated from 30-day rates) (%)
Acute PCI	Death	3.3	1.99	5.62
	NFMI	3.6	12.77	19.27
Acute CABG	Death	10.6	4.54	8.14
	NFMI	6.4	22.46	28.79
No acute revascularisation	Death	7.1	3.97	7.53
	NFMI	4.7	6.77	13.43

NFMI, non-fatal myocardial infarction.

TABLE 33 *Relative risk reductions relating to Strategy 1 used in sensitivity analysis 1*

Revascularisation group	Event	Base-case RR	Sensitivity analysis: separate RR for acute PCI/no acute PCI
Acute PCI	Death	0.84 (0.71–0.98)	0.83 (0.53–1.29)
	NFMI	0.94 (0.87–1.02)	0.80 (0.65–0.95)
Acute CABG	Death	0.84 (0.71–0.98)	0.91 (0.81–1.04)
	NFMI	0.94 (0.87–1.02)	0.95 (0.86–1.03)
No acute revascularisation	Death	0.84 (0.71–0.98)	0.91 (0.81–1.04)
	NFMI	0.94 (0.87–1.02)	0.95 (0.86–1.03)

RR, relative risk.

applies the relative risks reported in Boersma and colleagues' paper,[265] which relates to Strategy 1 trials only. The reason for this is that Boersma and colleagues' meta-analysis was able to estimate separate relative risks for those patients undergoing/not undergoing acute PCI (i.e. PCI within 5 days) and the baseline probabilities of an acute PCI varies between UK data sources and the trials (see above). *Table 33* provides details of the relative risks used in the two separate analyses for Strategy 1.

Table 34 shows that neither of the two additional sensitivity analyses using the revised baseline event data results in a change of the relative ordering of the strategies in terms of mean costs and QALYs. As before, in each of the analyses, Strategy 2 is dominated and Strategy 3 is ruled out because of extended dominance. The impact of changing the baseline event rates from UK-specific to trial sources, but not the relative risks, reduces the ICER of Strategy 1 from £5738 to £5753. The slight increase in uncertainty surrounding this decision is reflected in the lower probability that Strategy 1 is cost-effective in comparison to the base-case estimates. Although the revised baseline event rates have minimal impact on the ICER of Strategy 1 and do not appear to alter the optimal adoption decision, they do have a significant impact on the comparison between Strategies 3 and 4. The ICER of Strategy 3 relative to Strategy 4 falls from £25,811 in the base-case model to £11,160 using the revised baseline event data. However, Strategy 3 is still ruled out by Strategy 1 on the basis of extended dominance.

The impact of changing both the baseline event rates and using separate relative risks for acute PCI/no acute PCI for Strategy 1 has a greater impact on the results (*Table 34*). The ICER for Strategy 1 increases from £5667 to £9609. There is also greater uncertainty associated with the optimal decision. Since the revised assumptions for this sensitivity analysis only alter the relative risks applied to Strategy 1, the ICER for Strategy 3 in comparison with Strategy 4 remains the same.

TABLE 34 *Results of sensitivity analyses using baseline data event data from the trials rather than UK-specific sources*

Sensitivity analysis	Strategy	Cost (£)	QALY	ICER	Probability cost-effective for maximum WTP		
					£10,000	£30,000	£50,000
(a) Same RR as applied in base-case model	1	14,235	7.7921	£5753	78.97	91.71	93.08
	2	13,787	7.6728	D	2.56	1.43	1.23
	3	13,844	7.7101	ED (£11,160)	6.82	4.88	4.35
	4	13,678	7.6952		11.65	1.98	1.34
(b) Strategy 1 RR based on patient-level meta-analysis.[265] Differential RR applied to acute PCI/no acute	1	14,174	7.7468	£9609	45.30	70.03	73.92
	2	13,787	7.6728	D	6.45	4.8	4.22
	3	13,844	7.7101	ED (£11,160)	18.51	16.92	15.77
	4	13,678	7.6952		29.74	8.25	6.09

WTP, willingness to pay.

However, as in the previous sensitivity analysis, Strategy 3 is still ruled out by Strategy 1 by extended dominance.

Sensitivity analysis II: increased use of clopidogrel

The second sensitivity analysis assessed whether likely changes in UK practice involving use of clopidogrel would impact on the cost-effectiveness of GPAs. The relative risks applied in the model based on the CURE trial[273] were as follows: all-cause death 0.92 (95% CI 0.79 to 1.05); non-fatal MI 0.77 (95% CI 0.67 to 0.89); all revascularisation 0.92 (95% CI 0.85 to 0.98); and major bleeding 1.38 (95% CI 1.13 to 1.67). The results of this analysis are shown in *Table 35*. Using the same assumptions as applied in the base-case model, clopidogrel is ruled out through extended dominance by Strategy 1. However, when the more conservative relative risk estimates derived from Boersma and colleagues' patient-level meta-analysis[265] are applied to Strategy 1, clopidogrel now appears to be the optimal strategy, ruling out Strategies 1 and 2 by dominance and Strategy 3 by extended dominance. The resulting ICER for clopidogrel in comparison to Strategy 4 is £6978.

Discussion

The purpose of this case study has been to show some of the methods that can be employed to make the results of decision analytic cost-effectiveness models as specific as possible to a particular decision-maker or jurisdiction. The case study presented here relates to a decision analysis of the cost-effectiveness of GPAs, where the jurisdiction was the UK NHS and the decision-maker was NICE. The starting point of the case study is that the purpose of decision models in this context is to help decision-makers identify the most cost-effective intervention for a given group of patients. Models are of particular value when a range of sources of data inputs need to be brought to bear on a decision problem and when those data inputs are subject to uncertainty. In practice, this is likely to be the case with the majority of decision problems.[12] Furthermore, the expectation with decision models is that part of the process of defining a decision problem is to specify the decision-maker or jurisdiction on whom the analysis is targeted. This requirement to make models specific to jurisdiction requires careful selection of model structure and data inputs. In many cases, it will be necessary to adapt parameter estimates identified in the literature to make them specific to a jurisdiction.

As the review of applied models in osteoporosis in Chapter 7 reveals, analysts will invariably ensure that cost inputs into models are specific to the target decision-maker. This is paralleled in this case study with the use of UK-specific data on resource use and unit costs. However, it is typically the case that models assume that effectiveness data are transportable across locations – particularly between countries. This case study has taken the example of the management of ischaemic heart disease where UK practice is known to differ from that in other developed countries, particularly with respect to revascularisation rates.[267] It was, therefore, inappropriate to use the absolute effectiveness measures estimates in relevant clinical trials. Rather, the model was populated by

TABLE 35 *Results of sensitivity analysis including clopidogrel as a fifth management strategy*

					Probability cost-effective for maximum WTP[a]		
Sensitivity analysis	**Strategy**	**Cost (£)**	**QALY**	**ICER**	**£10,000**	**£30,000**	**£50,000**
(a) Add clopidogrel as a fifth option	1	12,723	7.7862	£5750	61.04	71.27	72.32
	5 (clopidogrel)	12,526	7.7405	ED (£6978)[b]	30.85	27.32	26.67
	2	12,244	7.6825	D	0.12	0.12	0.13
	3	12,223	7.6896	ED (£26,296)[c]	0.29	0.48	0.56
	4	12,152	7.6869		7.7	0.81	0.32
(b) Add clopidogrel as a fifth option plus use patient-level meta-analysis for RR for Strategy 1	1	12,684	7.7438	D	26.76	42.19	44.25
	5 (clopidogrel)	12,526	7.7457	£6978	52.04	52.87	52.27
	2	12,244	7.6879	D	0.44	0.34	0.41
	3	12,223	7.6946	ED (£26,296)[c]	0.86	1.95	1.84
	4	12,152	7.6923		19.9	2.65	1.23

[a] The probability that each strategy is more cost-effective than the others conditional on different maximum willingness to pay for an additional QALY.
[b] ICER clopidogrel versus Strategy 4.
[c] ICER strategy 3 versus Strategy 4.

separating the absolute effects into baseline event rates and a relative treatment effect. Given the differences in UK practice with respect to the management of non-ST-elevation ACS, it was assumed that the baseline data from the non-UK trials would not be directly relevant to the NHS. Instead, baseline data were taken from available observational sources, in particular PRAIS-UK.[269] However, unbiased estimates of relative treatment effect (i.e. relative risk) rely on experimental design, so available trial evidence is always likely to be the main source of this information.

In the case study, therefore, a random effects meta-analysis was employed to pool the estimates of relative treatment effect. However, as for more general methods for assessing the generalisability in clinical results between patient subgroups,[19] the assumption of independence between baseline and relative risks should be tested where possible. This was explored in this case study using meta-regression. No strong statistical relationship was identified in these analyses, but the successful implementation of the approach relies on there being a reasonable number of trials which represent the 'observations' in the analysis. In the case of Strategy 1 trials in the case study, only seven trials were available and for Strategy 3 six trials. Fitting a line between a small number of points in this way is likely to be unreliable. The case study model therefore assumes an independence between baseline risks (taken from UK-specific sources) and relative treatment effect (taken from the international trials).

One of the implications of the methods presented in the case study is the potential importance of observational data in assessing the cost-effectiveness of healthcare technologies. These sources are valuable not for their estimate of relative treatment effect, but to provide information on baseline event rates and also relating to long-term extrapolation.

The process of decision modelling involves the synthesis of available data in the knowledge that decisions have to be taken (implicitly or explicitly) regardless of the quality of the data inputs.[12] In seeking data inputs which are specific to the target jurisdiction, there is also a need to judge how adequately those inputs reflect the trend in practice in the relevant location given that published data are likely to reflect clinical practice in the past and this may be consistent with present practice. In the case study here, sensitivity analysis was used to assess the implications of two trends in clinical practice in the UK which were probably not reflected in the observational data: the increasing use of PCI and the use of clopidogrel as another management option in non-ST-elevation ACS. The results of the case study suggest that the first of these has little impact of the optimal strategy but, under some conditions, the latter will.

In many instances it will not be possible to identify data sources which relate directly to the target jurisdiction, and using estimates from other locations will be the only option. In these circumstances, there are several possible

approaches. One is to use expert opinion and information from other disease areas and technologies to assess the extent to which a particular parameter is likely to vary between locations. Using this information, adjustments could be made to the mean estimates of that parameter in the model or to the uncertainty around that mean represented by its distribution. An alternative is to assess how important the particular parameter is to the decision about the optimal form of management. This can be undertaken using threshold analysis,[16] which asks how much that parameter needs to change in order for the optimal decision to alter and a judgement can then be made regarding the likelihood that the parameter would take that 'threshold value'.

Another approach is to use formal value of information analysis,[153] where the contribution of the uncertainty in the parameter of interest to the overall uncertainty in the optimal decision is quantified and valued in monetary terms. The advantage of this method is that it can potentially indicate how sensitive the optimal decision is to whether jurisdiction-specific data are used and also show the potential value of additional research to estimate a particular parameter for a specific jurisdiction. In short, the importance of the absence of appropriate data which are taken directly from the target jurisdiction to a particular analysis is an empirical issue and depends on the sensitivity of the choice of optimal strategy to the uncertainty in that parameter.

Conclusions

This case study has sought to illustrate how decision analytic modelling can be used to pool available evidence to inform particular decision-makers about the most appropriate allocation of resources relating to a particular healthcare intervention. Part of the modelling process is to establish how appropriate the decisions about model structure, the assumptions which underly it and the data inputs are to the particular decision-maker and jurisdiction. This process will often require additional analysis and data to adjust available estimates.

Chapter 9
Conclusions

Summary of findings

Variability in cost-effectiveness by time and place

Several studies were identified that considered factors that may generate variability in cost and/or cost-effectiveness between settings (Chapter 3). The most cited factor in the literature is the unit costs associated with particular resources. Both between centres (within country) and between countries, it would be expected that variations may exist in the unit cost of resources such as hospital stay, clinical staff and outpatient attendances. A number of factors cited in the literature as independent sources of variation in economic results may also partly determine differences in unit costs (e.g. economies of scale and case mix). Some of the most frequently cited factors are as much associated with the measurement of **effectiveness** as with cost-effectiveness. These include the artificial characteristics of centres undertaking research (e.g. randomised trials), patient case mix and clinical practice variation. No studies were identified which considered explicitly possible factors causing variability in the results of economic studies over time. However, there is a strong degree of consistency in the principles that can explain variability in costs and effectiveness between locations and across time.

A large number of studies have been identified which have sought to quantify variability using either trials or decision models as the 'vehicles' for the evaluation. These include studies which have taken an existing trial or model focused on a specific location and incorporated a different set of unit costs. Some studies have collected resource use data in a number of centres and countries (often from clinical experts rather than direct observation) and then explored variation in those measures between locations. Not surprisingly given the large number of factors which can, in principle, lead to variation in economic variables, several authors have shown important differences in the volume and cost of resource use between locations. Relatively few studies have considered variability in outcomes as assiduously as costs, reinforcing the view that most authors implicitly consider that clinical effectiveness measures are more exchangeable across locations than cost data.

Again, few studies were identified looking at variability in the results of economic evaluation over time. Indeed, only one study[107] was identified which explicitly considered this type of variation (since the review in this report was undertaken, a paper has been published which considers a range of issues regarding variability in cost-effectiveness over time[108]).

Methods to assess variability in cost-effectiveness by time and place

The methods literature in economic evaluation contains a range of approaches to quantifying variation by time and place (Chapter 4). In the context of studies based on patient-level data (e.g. randomised trials) undertaken in single locations, one way in which this issue has been addressed is in terms of the design and reporting of studies (e.g. more careful thought about the selection of the location for trials). Although design and reporting of these single-location studies is a necessary part of assessing likely generalisability, it is unlikely to be sufficient. In many cases, there will be good reasons to think that factors such as patient mix, resource use or unit costs are different in other locations to that chosen for the primary study. The decision analytic model is the main means by which this sort of adaptation has been undertaken. Usually, the focus is taking the results of a trial undertaken in one country and extrapolating to another, but the principles apply also to extrapolation between centres within a single country. Although most of the extrapolation models reviewed in this chapter focus on changes to the cost side of the evaluation, it is clear that there is a range of reasons why the effect side may also need to be adapted from one setting to another. Methods identified in this review involve the view that baseline event rates in studies tend to be location-specific, whereas the relative treatment effect is more exchangeable across subgroups and locations. It is important to note, of course, that the assumption about the exchangeability of relative treatment effects should, as far as possible, be tested.

The multinational trial has emerged as an important vehicle for clinical evaluation in recent years, particularly for pharmaceuticals. An important premise of such trials is that the

treatment effects of the technologies being evaluated will be fairly generalisable across locations. There has also been a growth of multicentre trials within single countries, again in an attempt to recruit sufficient patients most rapidly. The starting point for the papers reviewed here is that variables within an economic evaluation are likely to vary between locations. However, a further factor is that basic economic theory would suggest that these local variables are related through the production function. The challenge for economic evaluation alongside multilocation trials is how best to reflect this interdependence between costs, resource use and outcomes and the likely variation between locations. Much of the literature ignores this interdependence by applying unit costs from one or a few centres/countries to pooled resource use, and to relate to pooled outcome data. An emerging methods literature is seeking appropriate forms of analysis (usually based on regression models) which accept that some components of resource use and outcomes are exchangeable across locations whereas some are not.

There will always be a need to extrapolate cost-effectiveness results from multilocation trials to one or more locations which did not recruit patients. Future development of regression methods may allow this to be undertaken more systematically using covariates to characterise centres. There is also a key role for the decision analytic model as a vehicle for adjusting the cost-effectiveness estimates in primary studies, such as randomised trials, to locations which did not recruit to those studies. When deciding which data to use to populate a model, the extent to which they are directly applicable to the particular decision-maker who is the target of the analysis needs to be considered. As discussed above, adjustments will be necessary if data are not considered exchangeable between locations.

The review failed to identify a major literature on variability in cost-effectiveness over time. There is a growing literature, however, on the use of iterative methods to evaluate healthcare technologies,[28,150] which was considered outside the scope of this review. These methods are based on the premise that the cost-effectiveness of a technology will vary over time as new research data emerge about its use. More fundamentally, formal economic analysis and statistical decision theory is used to identify the priorities for additional research.[151–153] A dynamic process emerges whereby a common decision analytic framework is used to inform decision-making about the use of particular technologies based on existing information and about the needs for future research and its optimal design.

Dealing with variability by location in economic studies alongside multicentre trials

The review of applied economic evaluations alongside multilocation trials identified some important features of the methods which are **actually** being used to assess variability by location (Chapter 5). This provides a useful counterpoint to the methods literature summarised above and the case studies. The purpose of the review was not to judge whether the studies in the review were generalisable in themselves, but rather to establish whether enough information was provided for a given decision-maker to assess their relevance to their own jurisdiction. The review indicated that some of the general features of good reporting are not being widely adhered to. The study perspective was defined in only 42 studies (42%). If analysts neglect to provide this sort of information, the onus is on the decision-maker to interpret the perspective from available information on costs and effects. A further surprising result was that only 25 studies (25%) reported resource use and unit costs separately, with the rest of the sample reporting them partially separate ($n = 31$) or combined ($n = 45$). The lack of separate reporting prevents the decision-maker from assessing the extent to which the vector of unit costs used in a particular study is relevant to his/her jurisdiction.

Some other aspects of the observed reporting provide further barriers to the assessment of a study's generalisability and relevance to particular jurisdictions, but these have not been so widely reflected in general reporting guidelines for economic evaluation. First, only two studies assessed the representativeness of the study's sample with respect to the study population in the trial centres. Second, in 91 articles, a definition of the study setting was provided, but only three papers described the characteristics of the healthcare system(s) where the economic analysis was conducted. In addition, only seven studies reported one or more centre-specific characteristics such as the volume of cases they treat.

The sample of applied studies alongside multilocation trials showed little use of the statistical approaches identified in the methods review to assess variability by location. There may be a number of reasons for this. First, a large

proportion of the studies included in this review were published between 1995 and 1998 and, at that point, some of the methodological contributions in this field were still to be published. Second, it is likely that editorial needs may have limited the scope for exploratory work within the main economic paper. Finally, the focus of the paper may have been clinical and the economic analysis was a secondary consideration. The review found no studies which had used analytical methods to extend the within-trial results to other locations. The review did show the use of sensitivity analysis to assess a range of different uncertainties in trial-based studies. Although these rarely set out to assess variability between locations, decision-makers may be able to interpret them in this manner.

In the case studies reported in Chapter 6, the regression-based approaches to quantifying variability by location identified in Chapter 4[25,135] have been taken a step further using MLM. The approach formally segments variation in net benefit into that occurring at the level of individual patients and that at the level(s) of locations such as centre and country. Where cost and/or outcome data are clustered by location, these methods facilitate correct estimates of the uncertainty in cost-effectiveness results. MLM also provides a means of estimating location-specific cost-effectiveness. Here 'location' could mean centres (e.g. hospitals) within a jurisdiction or countries. In the past, this has been attempted in various ways, including the application of location-specific unit costs to all resource use and outcome data, and the analysis of data only relating to patients randomised in the location of interest (see Chapter 5). The MLM framework uses empirical Bayes estimation to predict location-specific estimates of cost-effectiveness. As part of this, location-specific estimates are shrunk back to the overall mean, with the degree of shrinkage depending on the sample size in a centre and the level of variability in net benefits between and within centres. In other words, using these methods, a location-specific estimate of cost-effectiveness is a combination of the overall mean across patients in all location, and the mean in patients within that location.

The extent to which the use of MLM is crucial in a particular study depends on the proportion of overall variability in net benefit that takes place between locations. Although, with relatively small values of the intra-class correlation coefficient (*ICC*), reasonably good agreement between multilevel models and standard methods (using or equivalent to OLS regression) can be expected, in practical terms it is impossible to establish a rigid threshold value of the *ICC* above which the use of MLM should be recommended. There is a strong basis for starting the analysis with a multilevel regression to understand the data structure in more depth. However, in two of the three case studies considered in Chapter 6, it has not yet been possible to identify an appropriate specification of an MLM, so additional research is required in this area.

An important policy issue is raised by this work: the extent to which location-specific estimates of incremental net benefit are useful to decision-makers. In the context of multinational trials, the ability to generate estimates of cost-effectiveness by country would seem potentially useful to country-level decision-makers. In the case of multicentre trials in a single country, however, this may not be so straightforward. The implication of generating centre-specific estimates is that the decision-maker may be willing to fund a particular intervention in one centre but not another. Alternatively, the analysis might be useful to centre-level decision-makers. The use of location-specific covariates in MLM to explain variation in cost-effectiveness between centres and countries may have more direct policy relevance. For example, an intervention might be cost-effective in centres with high levels of patient throughput but not in centres treating fewer patients. This might provoke policy makers to reassess the appropriate organisation of services. However, the relevance, in policy terms, of centre-level estimates of cost-effectiveness requires more exploration.

Although the implications of this need further consideration, it remains the case, in principle, that the multilevel population average from a random slope model remains the most appropriate way of estimating average cost-effectiveness (i.e. across centres), given clustered data, if the decision-maker is not interested in implementing different decisions in different centres.

Use of decision analytic models to provide location-specific estimates of cost-effectiveness

Chapter 7 detailed a review of decision analytic cost-effectiveness models applied to the area of osteoporosis. The aim of this chapter was to see whether analysts had identified a target decision-maker or jurisdiction, whether their data were appropriate for that decision-maker and what methods had been used to adapt data from one

location to another. The purpose of cost-effectiveness decision models is to aid decision-makers regarding appropriate resource allocation given available data. As such, it would seem logical for any analysis to be clear about the jurisdiction and specific decision-maker for which the model is designed. However, only three (16%) of the studies in this review provided these details explicitly, although it was possible to infer the decision context from other information provided. Similarly, authors frequently report the methods and results of studies for which a firm research question has not been stated.

Once the target decision-maker/jurisdiction and decision problem have been established, the former will need to decide whether the data inputs and assumptions in the model are the best available for their context. The majority of the reports provided sources for clinical and economic data and their valuation. The data inputs to those models for which a decision-maker was stated or could be inferred appeared to be relevant and, as far as could be judged from only the study report, 'best available' for the decision context. The studies in the review generally made more effort to ensure (and to be seen to ensure) that their cost inputs (at least unit costs) were specific to their target jurisdiction. Most papers were prepared to use clinical data from studies undertaken outside their context. As in other clinical areas (see Chapter 4), an exception to this assumption of the transferability of clinical data is the adjustment of baseline risks to make them specific to a particular jurisdiction (usually country) whilst assuming that the **relative** treatment effect is exchangeable geographically. This adjustment of baseline risks was undertaken rarely in the sample of papers reviewed.

Little attempt was made to justify the particular utilities used in those cost–utility models in the sample with respect to the target decision-maker. In part, this probably reflects the limited amount of utility data available relating to osteoporosis. Increasingly, decision-makers will be specific about the type of health state utility data that they wish to see in economic evaluations.

One aspect of the review was to assess the extent to which studies had assessed the impact of variability in parameter estimates associated with location using sensitivity analysis. In principle, this sort of analysis might be undertaken for two reasons. First, there may be variability in clinical and/or cost parameters **within** a given jurisdiction. If there was good reason to think that this level of variability might impact on the conclusions of the analysis, sensitivity analysis would be strongly indicated. The second way in which sensitivity analysis could be used is to assess whether the results of the model, as they apply to the target jurisdiction, would also apply to other locations by appropriate variation in parameter values.

If it is accepted that the purpose of decision models is to address decision problems for particular jurisdiction/decision-makers, then this process of generalisation should probably not be seen as an essential element of model-based economic analysis. Although most of the studies in the review undertake extensive sensitivity analysis, few explicitly do this in order to explore variation between locations/jurisdictions. Only one of the models used probabilistic sensitivity analysis and most of the standard sensitivity analysis was undertaken to assess the importance of parameter uncertainty to the results, rather than variability between locations. The failure to assess variability within and between locations may have reflected the view that parameter uncertainty dwarfed variability within the target jurisdiction and generalising their results across locations was not a primary concern.

The purpose of the case study in Chapter 8 was to show some of the methods that can be employed to make the results of decision analytic cost-effectiveness models as specific as possible to a particular decision-maker or jurisdiction. The case study related to the cost-effectiveness of GPAs for non-ST-elevation acute coronary syndrome, where the jurisdiction was the UK NHS and the decision-maker was NICE.

As noted above, within a modelling framework, cost data are usually tailored to the targeted jurisdiction, but effectiveness data are often assumed to be transportable across locations – particularly between countries. The case study has taken the example of the management of ischaemic heart disease where UK practice is known to differ markedly from that in other developed countries, particularly with respect to revascularisation rates. The model was populated by separating the absolute effects into baseline event rates and a relative treatment effect. Given the differences in UK practice with respect to the management of the disease, it was expected that the baseline data from the non-UK trials would not be directly relevant to the NHS. Instead, baseline data were taken from available observational sources, However, unbiased estimates of relative treatment effect (i.e. relative

risk) rely on experimental design, so available trial evidence is always likely to be the main source of this information.

In the case study, therefore, a random effects meta-analysis was employed to pool the estimates of relative treatment effect. However, the assumption of independence between baseline and relative risks should be tested where possible. This was explored in this case study using meta-regression. No strong statistical relationship was identified in these analyses, but the successful implementation of the approach relies on there being a reasonable number of trials which represent the 'observations' in the analysis. The case study model therefore assumed an independence between baseline risks (taken from UK-specific sources) and relative treatment effect (taken from the international trials). One of the implications of the methods presented in the case study is the potential importance of observational data in assessing the cost-effectiveness of healthcare technologies. These sources are valuable not for their estimate of relative treatment effect, but to provide information on baseline event rates, and also relating to long-term extrapolation.

The process of decision modelling involves the synthesis of available data in the knowledge that decisions have to be taken (implicitly or explicitly) regardless of the quality of the data inputs.[12] In seeking data inputs which are specific to the target jurisdiction, there is also a need to judge how adequately those inputs reflect the trend in practice in the relevant location given that published data are likely to reflect clinical practice in the past and this may be inconsistent with present practice. In the case study, sensitivity analysis was used to assess the implications of two trends in clinical practice in the UK which were probably not reflected in the observational data: the increasing use of PCI and the use of clopidogrel as another management option in non-ST-elevation ACS. The results of the case study suggest that the first of these has little impact on the optimal strategy but, under some conditions, the latter will.

In many instances, it will not be possible to identify data sources which relate directly to the target jurisdiction and using estimates from other locations will be the only option. In these circumstances, there are several possible approaches. One is to use expert opinion, or information from other disease areas and technologies, to assess the extent to which a particular parameter is likely to vary between locations. Using this information, adjustments could be made to the mean estimates of that parameter in the model or to the uncertainty around that mean represented by its distribution. An alternative is to assess how important the particular parameter is to the decision about the optimal form of management. Another approach is to use the formal value of information analysis,[153] where the contribution of the uncertainty in the parameter of interest to the overall uncertainty in the optimal decision is quantified and valued in monetary terms. The advantage of this method is that it can potentially indicate how sensitive the optimal decision is to whether jurisdiction-specific data are used, and also show the potential value of additional research to estimate a particular parameter for a specific jurisdiction. In short, the importance of the absence of appropriate data taken directly from the target jurisdiction to a particular analysis is an empirical issue, and depends on the sensitivity of the choice of optimal strategy to the uncertainty in that parameter.

Recommendations and future research

In this section, some of the implications of the research project are identified, distinguishing those that are relevant to researchers, decision-makers and research funders. These implications are considered separately for economic studies based on patient-level datasets (e.g. randomised trials) and for decision analytic models. Where it is considered appropriate, recommendations for future research are also included here.

Economic evaluation using patient-level data

The economic analyst has three opportunities to increase the generalisability of his or her study. First, at the **design** stage, the need for generalisability of findings can be anticipated. Second, in the **analysis** of results, qualitative and quantitative approaches can be used to produce findings relevant to a range of settings. Finally, in the **reporting** of results, attempts can be made to accommodate the needs of users/decision-makers in different geographical locations. This section discusses these issues in the context of patient-level data, with a focus on trial-based studies, and makes several recommendations for improved practice. In doing so, it draws on the review of trial-based studies (Chapter 5) and the case study on multilevel modelling (Chapter 6).

Recommendations in relation to design, analysis and reporting are given separately below, whilst recognising that these issues are often interlinked.

Recommendations in relation to design of studies

Clinical trials are primarily designed to estimate clinical parameters, for which generalisability has traditionally been considered less of an issue. Therefore, the economist seeking design changes for the purposes of increased generalisability will probably have to compromise on the range of changes sought. Nevertheless, several changes are possible and it may be argued that some serve the clinical objectives of the trial, in addition to the economist's need for increased generalisability of findings (e.g. the selection of a comparator that is widely used in a number of countries). Much depends on the body funding the study and its objectives. For example, a pharmaceutical company may be seeking to appeal to several jurisdictions, whereas NHS R&D may only be interested in implications for the NHS.

Selection of study sites (i.e. centres)

- For purposes of generalisability, selection of study sites would ideally focus on those that are representative of the jurisdiction(s) for which economic data are required. In principle, this could be a single site, but is more likely to be several sites, in order to reflect the variation in healthcare provision within and between different jurisdictions.
- If the intention is to apply multilevel modelling techniques in the analysis of the economic data, it would be useful to collect data on centre characteristics that could be used as covariates in the multilevel model. The same would apply to jurisdiction characteristics if the trial were being performed in more than one jurisdiction. These covariates will increase the efficiency of average cost-effectiveness estimates and, by looking at interactions with treatment, facilitate subgroup analysis by location characteristics.
- A statistical analysis plan should define covariates at all levels for which data are to be collected and proposed analytical methods should be clearly stated.
- More research is required to determine which characteristics are most useful as covariates in such a multilevel model, so initially it would be wise to collect a wide range of data. Some initial suggestions are given in *Table 36*.
- It would improve the efficiency of the model if the centres were selected randomly from the relevant population and a reasonable number (i.e. 15–20) were included in the trial. It would also be ideal to have a minimum number of observations (patients) in each centre, in order to make sure that the cluster characteristics are adequately represented.

Inclusion/exclusion of patients

- It is typical, in clinical trials, to have criteria for inclusion and exclusion of patients. In order for an economic evaluation based on the trial to be generalisable, the patients included should reflect the normal clinical caseload. Therefore, there would be concerns if a large percentage of patients were excluded from the trial.
- Another threat to generalisability would be present if the 'normal' caseload varies from place to place. This could arise if participating centres differ in respect of their catchment populations. In such situations it would be important to have a wide range of centres in the trial.
- It would also be important to collect a number of patient-level variables that could be used as covariates in a multilevel model. These covariates could include age, gender, socio-economic status and previous medical history. This would facilitate more efficient estimates of average treatment cost-effectiveness and allow the estimation of cost-effectiveness relating to subgroups based on patient characteristics.[156] Patient-level variables are typically collected in trials already, in order to check for imbalance at

TABLE 36 *Possible higher-level covariates to use in a multilevel model*

Jurisdiction level	Centre level[a]
% of GDP spent on healthcare Reimbursement system for hospitals Payment method for physicians	Volume of relevant cases treated Bed occupancy Hospital type (e.g. teaching or non-teaching) Range of clinical specialties Geographical location (e.g. large city or small town)

GDP, gross domestic product.
[a] These covariates would be relevant if the 'centre' were a hospital. However, in principle, the 'centre' could be a region, health authority or individual clinician.

baseline, so this should not impose any additional data collection burden.
- There may also be instances where centres' characteristics determine their typical caseload. For example, centres of excellence typically treat more serious cases than normal general hospitals. In these situations there is likely to be an interrelationship between centre characteristics and patient characteristics. Hence it becomes crucial to collect both patient-level and centre-level variables.

Selection of comparator therapy
- The comparator selected needs to be relevant to the jurisdictions in which the study is going to be used. Therefore, a threat to generalisability could exist if 'current practice' varies from place to place.
- In some cases it may be possible to agree on one or more compromise comparator(s) which reflect(s) normal practice in a wide range of settings.
- The alternative approach would be to let the clinician or centre select their own comparator therapy.[278] In this case it would be important to ensure that the trial includes a representative sample of centres and/or physicians, and again there is value in using MLM to explore variation in cost-effectiveness by location.

Perspective of the study
- The various international guidelines for economic evaluation have differences with respect to study perspective. Some recommend adopting a societal perspective, whereas others focus on government expenditure or a particular budget (e.g. the drugs budget). Therefore, the recommended approach, bearing in mind the need for generalisability, would be to adopt a broad societal perspective whilst retaining the capability to present costs and benefits by a range of different perspectives. There are also strong normative reasons for adopting the societal perspective.[279]

Collection of resource use and cost data
- The main recommendation here is to collect resource use data (e.g. hospital days, intensive care unit days, district nurse visits) separately from the unit costs or prices of those resources. The reasons for this are obvious. First, decision-makers considering a study undertaken in another location need to assess whether the practice patterns (and resulting resource use) observed in the study apply in their own setting. Second, decision-makers may wish to apply their own prices to the units of resource use.

Health state preference values
- Health state preference values can be obtained from the literature, estimated directly on patients in the trial, or derived by using a generic instrument (e.g. EQ-5D, Health Utilities Index). The generic instruments use a questionnaire, administered during the trial, to classify patients into health states. The set of values for states (i.e. the tariff) is then provided with the instrument, having been obtained from a community survey. For the purposes of generalisability, the health state valuations would ideally be relevant to the population(s) under study. For example, the NICE guidance to manufacturers and sponsors of health technologies states that "the most relevant values are those of the general population of England and Wales"[6] (p. 17). In most cases the use of a generic instrument is recommended, although the tariffs for EQ-5D and the Health Utilities Index are not available for many different populations.

Recommendations in relation to analysis of results
- Two approaches to the analysis of variability in cost-effectiveness by location, using data from multicentre or multinational trials, have been reported in the literature (these were reviewed in Chapter 4). First, Cook and colleagues[131] recommend a test of interaction approach, to explore the level of homogeneity in the data. This mirrors the approach frequently followed in the analysis of clinical data from multicentre trials. Namely, if no interaction exists between centre and treatment effect the data can be pooled, thereby giving a more precise estimate of treatment effect. Second, Willke and colleagues[25] used a fixed-effect regression approach, based on separate regressions for cost and outcomes, whereby country dummy variables are introduced alongside other explanatory variables. The further development proposed here is to use MLM (Chapter 6) and, although further methods research is needed to identify the best way of applying these methods, this approach should be considered as part of the analysis of multilocation trials.
- The advantage of the use of MLM is that, if patient-level data are clustered by location, it will provide more appropriate estimates of the uncertainty around an intervention's cost-effectiveness; it can also facilitate location-specific estimates of cost-effectiveness. At the least, the approach can be used to consider the degree of clustering in data and hence the extent to which this should be reflected in the full analysis.

- A number of further research issues arise in the context of MLM. These include the overall specification of the models; selection of patient- and location-level covariates and the specification of their interaction with treatment; the appropriate MLM approach when there are a number of levels in the data hierarchy (e.g. patients, surgeons, centres, countries); appropriate methods when there are few locations in the trial; and the use of Bayesian approaches to multilevel modelling.
- Although the greater use of formal statistical methods, such as MLM, is warranted in trial-based studies, there will remain an important role for sensitivity analysis in exploring the implications of variation in some parameters (e.g. unit costs and preference values).
- There has been some use of econometric methods, such as selection models and instrumental variables, to adjust observational datasets for selection bias,[144] and some consideration of those methods to increase the generalisability of randomised trials,[143] perhaps in the context of comprehensive cohort analysis.[20] Further research is justified in the principles and application of these methods.

Recommendations in relation to the reporting of results

- Even if it has not been possible to address fully all the issues of generalisability at the design or analysis stage, the needs of study users can still be partly accommodated during the reporting of results. The recommendations are summarised in *Table 37*. The general objective is to help the users of studies decide whether a given study is relevant to their own setting.
- Clearly, these additional reporting suggestions will be constrained by the limitations of space, particularly by journals. There is, therefore, an argument for greater use of more detailed technical reports to be made available as supporting documents, perhaps on journal websites.
- An important area of further research relates to the policy relevance of the location-specific estimates of cost-effectiveness which MLM facilitates. Although the value of these results may be clear for individual countries in a multinational trial, the decisions which might be taken given centre-specific estimates of cost-effectiveness are less obvious.

Economic evaluation using decision analytic modelling

As described in Chapters 7 and 8, the use of a single patient-level dataset, such as a randomised trial, as a vehicle for economic evaluation frequently has a number of limitations. These include the partial nature of the comparisons undertaken, short-term follow-up, use of intermediate rather than ultimate measures of health outcomes and unrepresentative patients, clinicians and locations. Given the increasing need for policy-relevant cost-effectiveness research to inform particular decisions about the funding and reimbursement of healthcare interventions, these shortcomings of trial-based analyses will need to be addressed. The decision model represents an important analytic framework to

TABLE 37 *Recommendations for reporting the results of economic clinical trials*

Study element	Reporting recommendations
Study sites (centres)	Describe the characteristics of the centres participating in the trial. If these are from different countries, also report the relevant features of the various healthcare systems
Patient enrolment	Report the types of patients excluded from the trials and the percentage of the normal caseload that these represent. Comparison with the relevant patient population outside the trial centres
Treatment alternatives	Describe the alternatives in detail, so that study users can assess the relevance to their own setting
Perspective(s)	Report costs and benefits by each relevant perspective
Resource use and costs	Report quantities separately from prices/unit costs
Health state preference values	Report the source of the values and any instrument used
Analysis of variability	Provide details of quantitative analysis of variability by location. Ideally, this will be based on statistical analysis (such as multilevel modelling), but should at least incorporate standard sensitivity analysis
Other analytical issues	Provide details on the extent of incomplete observations (i.e. missing and censored data) Detail the characteristics of patients with incomplete data Describe the methods used to address the problem

generate estimates of cost-effectiveness based on a synthesis of available data and the explicit representation of uncertainty.[12] The extent to which decision models can and should consider variability by location, and their use as a means of adjusting clinical and economic results between locations, were considered in Chapters 7 and 8. A number of recommendations for decision model-based economic evaluation and suggestions for further research flow from this. Again, these points are arranged under design, analysis and reporting.

Recommendations in relation to design of studies

- Given the focus on a **decision**, any analysis should be clear about two important features of the research. The first is the specification of the decision problem, that is, the explicit statement of the options the cost-effectiveness of which is being compared and the patient group(s) for which the options are relevant. This key feature of the design of a decision model is a feature of most general guidelines in the area.[280,281]
- The second important feature is less frequently identified in these guidelines, and it relates to the decision-maker(s) and jurisdiction(s) whose decision the model is designed to inform. In some cases, a specific decision-maker might be specified, such as NICE. For other decision models, a more general focus may be suitable, such as Primary Care Trusts in England and Wales.
- Once these features have been defined, an important next stage is to ensure that the overall analytical approach and structure are appropriate to the relevant decision-maker(s). This will rely on the latter having made a clear statement about factors such as the perspective of the analysis (e.g. health service or societal) and the relevant objective function (e.g. generic health gain such as quality-adjusted survival or disease-specific outcomes). Sometimes there will be a lack of clarity about these factors, or they will vary between decision-makers when the model is targeted on more than one. In these circumstances, there is value in adopting the broadest perspective and objective function which will allow the results to be presented in several different ways.

Recommendations in relation to analysis of results

- The data which are used to populate a decision model should be justified given the stated target decision-maker(s) or jurisdiction(s). This will apply not just to unit costs, but to resource use, effectiveness and preference value data.
- Where several appropriate sources of data exist for a particular parameter, these should be appropriately pooled in such a way that the uncertainty relating to their precision and their possible heterogeneity is reflected in the model. This will involve standard meta-analysis[282] or more advanced methods of multiparameter synthesis.[283,284]
- When sources of evidence are available from within the target jurisdiction **as well as from outside**, an important issue is whether and how the latter should be incorporated. Further research is needed to develop methods of evidence synthesis which combine data from a range of jurisdictions and allow for the additional uncertainty in this process.
- When only data from outside the target jurisdiction are available, it is important to assess whether these can be assumed to be exchangeable across locations. In both the clinical and economic evaluation fields, **relative** treatment effectiveness is often assumed to be exchangeable across locations and patient subgroups, whereas baseline event rates are not. Given available data, the reliability of this assumption can be assessed empirically (see Chapter 8).[271]
- For preference values, available evidence (Chapter 3) suggests little systematic variation between locations (e.g. countries) in mean values, indicating that location-specific estimates may not be essential. In the case of resource use and costs, it would be expected that location-specific data would be required, given their known variability (Chapter 3). Further research would be valuable to look at the issues around using the same approach for resource use, costs and preference values as for effectiveness data, that is, taking a relative treatment effect as exchangeable across locations and the baseline as location-specific. An important feature of such research would again be to reflect the uncertainty associated with the assumptions regarding the location-related exchangeability in the decision model.
- In any decision model, there will be a range of different types of uncertainty to deal with explicitly and to reflect in the overall results and interpretation of the analysis. In this process, it is important to distinguish parameter uncertainty, which relates to the imprecision with which a parameter is estimated due to there being a finite sample, from variability or heterogeneity which is concerned with how parameter estimates vary across contexts. These 'contexts' could be patient subgroups or, as is the focus here, locations. As suggested above,

parameter uncertainty may need to include the implications of taking data from sources other than the main jurisdiction of interest – further research will illuminate how this might be implemented. Probabilistic models, where data inputs are incorporated as random variables, are the appropriate means of handling parameter uncertainty.
- When a model is targeted at more than one decision-maker/jurisdiction, an important aspect of the analysis will be to assess the **variability** in results between locations. This is feasible, using sensitivity or scenario analysis, as long as alternative parameter estimates exist for individual locations. These methods will be important for both multinational analyses and multilocation studies within a given country.

Recommendations in relation to the reporting of results

- The level of detail and complexity involved with many decision models means that communicating all aspects of model structure, assumptions and data inputs can be a major task. Some general guidelines for this process have been published elsewhere.[280,281] As noted above, the more extensive use of technical reports, to support journal articles, is likely to be very important for comprehensive communication.
- A key feature of reporting models is to be able to establish that each parameter input is appropriate for target decision-maker(s)/ jurisdiction(s). This is part of the more general reporting task of justifying all assumptions and parameter values, but it is recommended that this is clearly related to the target customer.
- Again, explaining the methods which have been used to 'preanalyse' data inputs so they are suitable for incorporation into models (e.g. meta-analysis), is part of the general reporting process for decision models. This should include any preanalysis which was undertaken to adjust parameters estimated from the location in which they are measured to that which is relevant to the model.

Increasing the generalisability of economic evaluations

Based on the above, those commissioning, undertaking or using economic evaluations require a framework for increasing the generalisability of studies. Therefore, two checklists are presented in *Tables 38* and *39* for trial-base and modelling studies, respectively. If those planning new research studies (or those reviewing existing studies) apply the checklists, it is likely that, over time, more studies will produce generalisable results.

TABLE 38 *Checklist for assessing the generalisability of trial-based studies*

1. Are study sites representative of the jurisdiction(s) for which data are required?
2. Is it possible to select the sites (centres) at random?
3. Can data on centre characteristics be collected (e.g. bed occupancy levels)?
4. Does the trial include a high proportion of the normal clinical caseload?
5. Does the comparator therapy (to the technology of interest) represent current practice in the settings concerned?
6. Is a wide range of user perspectives represented in the study?
7. Are prices (unit costs) being collected separately from resource use data?
8. Is a widely used generic instrument being used for QoL (e.g. utility) measurement?
9. Can regression-based techniques be used to obtain centre-specific cost-effectiveness ratios?
10. Are the reporting recommendations in *Table 37* being followed?

TABLE 39 *Checklist for assessing the generalisability of modelling studies*

1. Are the decision problem, the relevant settings and audiences (i.e. decision-makers) clearly specified?
2. Does the overall analytical approach incorporate the relevant perspectives (e.g. health service or societal) and relevant objective functions (e.g. maximising health gain)?
3. Are the data used to populate the model relevant to the target audiences (i.e. decision-makers) and settings?
4. Where data from different sources are pooled, is this done in such a way that the uncertainty relating to their precision and possible heterogeneity is adequately reflected?
5. If data from other settings are used, have these been assessed for relevance in the settings of interest?
6. Is uncertainty (i.e. parameter uncertainty and heterogeneity) adequately reflected in the model?
7. Are results reported in such a way that allows the assessment of the appropriateness of each parameter input and each assumption in the target settings?

Relevance of findings to the NHS

The objective of conducting economic evaluations is to improve healthcare decision-making. For economic evaluations to improve decision-making they need to be relevant to the decision-maker's setting, in this case the NHS. The issue addressed by this report is that many economic evaluations are not generalisable to other settings.

The reviews of the literature contained in the report assessed the extent of the problem and outlined the main causes of lack of generalisability. The case studies explored approaches to increasing the generalisability of trial-based and modelling studies.

This concluding chapter summarises the findings and presents recommendations for the design, analysis and presentation of economic evaluations in the future. If these recommendations are followed, it is likely that the results of economic evaluations will be more relevant to NHS decision-makers in the future.

Summary of recommendations for further research

Drawing on the material in this chapter, it is possible to summarise some important areas for further research. As far as possible, these have been placed in priority order.

- The development of methods of evidence synthesis which model the exchangeability of data across locations and allow for the additional uncertainty in this process. These methods should relate to all parameters relevant to economic evaluation.
- Assessment of alternative approaches to specifying multilevel models to the analysis of cost-effectiveness data alongside multilocation randomised trials.
- Identification of a range of appropriate covariates relating to locations (e.g. hospitals) in multilevel models.
- Further assessment of the role of econometric methods (e.g. selection models) for CEA alongside observational datasets and to increase the generalisability of randomised trials.

Conclusions

A large number of factors are mentioned in the literature that might be expected to generate variation in the cost-effectiveness of healthcare interventions across locations. Several papers have demonstrated differences in the volume and cost of resource use between locations, but few studies have looked at variability in outcomes.

In applied trial-based cost-effectiveness studies, few studies provide sufficient evidence for decision-makers to establish the relevance or to adjust the results of the study to their location of interest. Very few studies were utilising statistical methods formally to assess the variability in results between locations. In applied economic studies based on decision models, most studies either stated their target decision-maker/jurisdiction or provided sufficient information from which this could be inferred. There was a greater tendency to ensure that cost inputs were specific to the target jurisdiction than clinical parameters.

Methods to assess generalisability and variability in economic evaluation studies have been discussed extensively in the literature relating to both trial-based and modelling studies. Regression-based methods are likely to offer a systematic approach to quantifying variability in patient-level data. In particular, MLM has the potential to facilitate estimates of cost-effectiveness, both of which reflect the variation in costs and outcomes between locations, and also enable the consistency of cost-effectiveness estimates between locations to be assessed directly. Decision analytic models will retain an important role in adapting the results of cost-effectiveness studies between locations.

Acknowledgements

The authors wish to thank members of the Expert Advisory Group: Professor Keith Abrams, Professor Mark Baker, Dr Andrew Briggs, Dr Karl Claxton, Professor Tony Culyer, Professor Bernie O'Brien and Dr Nigel Rice. They would also like to thank Jimmy Christie for his administrative assistance with the NHS Economic Evaluation Database.

John Nixon (Centre for Reviews and Dissemination, University of York) contributed to Chapter 5. Nigel Rice (Centre for Health Economics, University of York) and Andrew Briggs (Health Economic Research Centre, University of Oxford) Contributed to Chapter 6. Stephen Palmer, Zoe Philips and Laura Ginnelly (Centre for Health Economics, University of York), Keith Abrams (University of Leicester) and Mike Robinson (University of Leeds) contributed to Chapter 8.

This report was commissioned by the NHS R&D HTA Programme. The views expressed in this report are those of the authors and not necessarily those of the NHS R&D Programme. Any errors are the responsibility of the authors.

Contributions of the authors

Mark Sculpher (Professor of Health Economics) contributed to the design of the study, directed the project and drafted chapters of the report. Francis Pang (Research Fellow in Health Economics) undertook some of the literature reviews and drafted chapters of the report. Andrea Manca (Research Fellow in Health Economics) undertook some of the literature reviews, developed the work on multilevel modelling and drafted chapters in the report. Mike Drummond (Professor of Health Economics) contributed to the design of the study and drafted chapters of the report. Su Golder (Information Officer) undertook literature reviewing and drafted sections of the report. Hege Urdahl (Research Fellow in Health Economics) undertook one of the systematic reviews and drafted the relevant chapter. Linda Davies (Reader in Health Economics) contributed to the design of the study. Alison Eastwood (Senior Research Fellow in Systematic Review) contributed to the design of the study and advised on systematic review methods.

References

1. Drummond MF, O'Brien BJ, Stoddart G, Torrance G. *Methods for the economic evaluation of health care programmes*. 2nd ed. Oxford: Oxford University Press; 1997.
2. Pritchard C. *Trends in economic evaluation. OHE briefing no. 36*. London: Office of Health Economics; 1998.
3. Gold MR, Siegel JE, Russell LB, Weinstein MC. *Cost-effectiveness in health and medicine*. New York: Oxford University Press, 1996.
4. Duthie T, Trueman P, Chancellor J, Diez L. Research into the use of health economics in decision making in the United Kingdom – Phase II. Is health economics 'for good or evil'? *Health Policy* 1999;**46**:143–57.
5. Hjelmgren J, Berggren F, Andersson F. Health economic guidelines – similarities, differences and some implications. *Value Health* 2001;**4**:225–50.
6. National Institute for Clinical Excellence. *Technical guidance for manufacturers and sponsors on making a submission to a technology appraisal*. National Institute for Clinical Excellence; 2001. URL: http://www.nice.org.uk (cited 23 January 2003).
7. Commonwealth Department of Health Housing and Community Services. *Guidelines for the pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee*. Canberra: AGPS; 1992.
8. Ministry of Health. *Ontario guidelines for economic analysis of pharmaceutical products*. Ontario: Ministry of Health; 1994.
9. Langley P, Martin R. Managed care guidelines for the economic evaluation of pharmaceuticals. *Am J Manag Care* 1997;**3**:1013–21.
10. Briggs AH, Gray A. Handling uncertainty when performing economic evaluation of healthcare interventions. *Health Technol Assess* 1999;**3**(2).
11. Briggs AH, Goeree R, Blackhouse G, O'Brien BJ. Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease. *Med Decis Making* 2002;**22**:290–308.
12. Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute for Clinical Excellence. *Lancet* 2002; **360**:711–15.
13. Brazier J, Deverill M, Green C, Harper R, Booth A. A review of the use of health status measures in economic evaluation. *Health Technol Assess* 1999;**3**(9).
14. Sculpher MJ. The role and estimation of productivity costs in economic evaluation. In Drummond MF, McGuire A, editors. *Theory and practice of economic evaluation in health*. Oxford: Oxford University Press; 2001. pp. 94–112.
15. Sassi F, Archard L, Le Grand J. Equity and the economic evaluation of healthcare. *Health Technol Assess* 2001;**5**(3).
16. Briggs AH, Sculpher MJ, Buxton M. Uncertainty in the economic evaluation of health care technologies: the role of sensitivity analysis. *Health Econ* 1994;**3**:95–104.
17. Sackett DL, Haynes RB, Tugwell P. *Clinical epidemiology: a basic science for clinical medicine*. Toronto: Little, Brown; 1985.
18. Kuntz K, Weinstein M. *Modelling in economic evaluation*. Oxford: Oxford University Press; 2001.
19. O'Connell D, Glasziou P, Hill S, Sarunac J, Lowe J, Henry D. Results of clinical trials and systematic trials: to whom do they apply? In Stevens A, Abrams K, Brazier R, Fitzpatrick R, Lilford R, editors. *The advanced handbook of methods in evidence based healthcare*. London: Sage; 2001.
20. Olschewski M, Schumacher M, Davis KB. Analysis of randomised and nonrandomised patients in clinical trials using comprehensive cohort follow-up study design. *Control Clin Trials* 1992;**13**:226–39.
21. Rice N, Jones AM. Multilevel models and health economics. *Health Econ* 1997;**6**:561–75.
22. Heckman JJ. Sample selection bias as a specification error. *Econometrica* 1979;**47**:153–61.
23. Jones AM. Health econometrics. In Culyer AJ, Newhouse JP, editors. *Handbook of health economics*. Amsterdam: Elsevier; 2000. pp. 265–344.
24. Drummond MF, Bloom BS, Carrin G, Hiliman AL, Hutchings HC, Knill Jones RP, *et al*. Issues in the cross-national assessment of health technology. *Int J Technol Assess Health Care* 1992;**8**:671–82.
25. Willke RJ, Glick HA, Polsky D, Schulman KA. Estimating country-specific cost-effectiveness from multinational clinical trials. *Health Econ* 1998; **7**:481–93.
26. Heyland DK, Kernerman P, Gafni A, Cook DJ. Economic evaluations in the critical care literature:

do they help us improve the efficiency of our unit? *Crit Care Med* 1996;**24**:1591–8.

27. Mason J, Drummond M. The generalisability of pharmacoeconomic studies. *Pharmacoeconomics* 1997;**11**:503–14.

28. Sculpher MJ, Drummond MF, Buxton M. The iterative use of economic evaluation as part of the process of health technology assessment. *J Health Serv Res Policy* 1997;**2**:26–30.

29. Murray CJL, Evans DB, Acharya A, Baltussen RM. Development of WHO guidelines on generalized cost-effectiveness analysis. *Health Econ* 2000; **9**:235–51.

30. Bryan S, Brown J. Extrapolation of cost-effectiveness information to local settings. *J Health Serv Res Policy* 1998;**3**:108–12.

31. Drummond MF. Comparing cost-effectiveness across countries. The model of acid-related disease. *Pharmacoeconomics* 1994;**5**:60–7.

32. Drummond MF, Brandt A, Luce B, Rovira J. Standardising methodologies for economic evaluation in health care. *Int J Technol Assess Health Care* 1993;**9**:26–36.

33. Jonsson B, Weinstein MC. Economic evaluation alongside multinational clinical trials. Study considerations for GUSTO IIb. *Int J Technol Assess Health Care* 1997;**13**:49–58.

34. Schulman KA, Buxton M, Glick H, Sculpher M, Guzman G, Kong J, *et al*. Results of the economic evaluation of the FIRST study. A multinational prospective economic evaluation. *Int J Technol Assess Health Care* 1996;**12**:698–713.

35. Baltussen R, Leidl R, Ament A. The impact of age on cost-effectiveness ratios and its control in decision making. *Health Econ* 1996;**5**:227–39.

36. Stason WB. Cost-effectiveness analysis in health care: opportunities and challenges to international comparisons. In Lasser U, Roccella EJ, Rosenfeld JB, Wenzel A, editors. *Costs and benefits in health care and prevention: an international approach to priorities in medicine*. Berlin: Springer; 1990. pp. 20–6.

37. Baker AM, Goldberg A, Arnold RJ, Kaniecki DJ. Considerations in measuring resource use in clinical trials. *Drug Inf J* 1995;**29**:1421–8.

38. Baltussen R, Ament A, Leidl R. Making cost assessments based on RCTs more useful to decision-makers. *Health Policy* 1996;**37**:163–83.

39. Goodman CA, Mills AJ. The evidence base on the cost-effectiveness of malaria control measures in Africa. *Health Policy Plan* 1999;**14**:301–12.

40. Revicki DA, Frank L. Pharmacoeconomic evaluation in the real world: effectiveness versus efficacy studies. *Pharmacoeconomics* 1999;**15**:423–34.

41. Drummond MF. The future of pharmacoeconomics: bridging science and practice. *Clin Ther* 1996; **18**:969–78.

42. Bailey KR. Generalising the results of randomized clinical trials. *Control Clin Trials* 1994;**15**:15–23.

43. Rittenhouse BE. The relevance of searching for effects under a clinical-trial lamppost: a key issue. *Med Decis Making* 1995;**15**:348–57.

44. Hughes DA, Bagust A, Haycox A, Walley T. The impact of non-compliance on the cost-effectiveness of pharmaceuticals. A review of the literature. *Health Econ* 2001;**10**:601–15.

45. Bonsel GJ, Rutten FF, Uyl de Groot CA. Economic evaluation alongside cancer trials: methodological and practical aspects. *Eur J Cancer* 1993;**29A** (Suppl 7):S10–14.

46. Coyle D, Davies L, Drummond MF. Trials and tribulations. Emerging issues in designing economic evaluations alongside clinical trials. *Int J Technol Assess Health Care* 1998;**14**:135–44.

47. Tolley K, Gyldmark M. The treatment and care costs of people with HIV infection or AIDS: development of a standardised cost framework for Europe. *Health Policy* 1993;**24**:55–70.

48. Wennberg J. Dealing with medical practice variations: a proposal for action. *Health Aff* (*Millwood*) 1984;**3**:6–32.

49. Ferguson B, Sheldon T, Posnett J. *Concentration and choice*. London: Pearson Professional; 1997.

50. O'Brien BJ. A tale of two (or more) cities: geographic transferability of pharmacoeconomic data. *Am J Manag Care* 1997;**3** (Suppl 5):S33–9.

51. Greiner W, Schoffski O, Graf VD, Schulenberg J-M. The transferability of international economic health results to national study questions. *Health Econ Prev Care* 2000;**1**:94–102.

52. Evans RG. *Strained mercy: the economics of Canadian health care*. Toronto: Butterworth; 1984.

53. Goeree R, Gafni A, Hannah M, Myhr T, Blackhouse G. Hospital selection for unit cost estimates in multicentre economic evaluations: does the choice of hospitals make a difference? *Pharmacoeconomics* 1999;**15**:561–72.

54. Mason J, Drummond MF, Torrance GW. Some of the guidelines on the use of cost effectiveness league tables. *BMJ* 1993;**306**:570–2.

55. Bennett CL, Armitage JL, Buchner D, Gulati S. Economic analysis in phase III clinical cancer trials. *Cancer Invest* 1994;**12**:336–42.

56. Bennett CL, George SL, Vose JM, Nemunaitis JJ, Armitage JL, Armitage JO, *et al*. Granulocyte-macrophage colony-stimulating factor as adjunct therapy in relapsed lymphoid malignancy:

implications for economic analyses of phase III clinical trials. *Stem Cells* 1995;**13**:414–20.

57. Haycox A, Drummond M, Walley T. Pharmacoeconomics: integrating economic evaluation into clinical trials. *Br J Clin Pharmacol* 1997;**43**:559–62.

58. Rizzo JD, Powe NR. Methodological hurdles in conducting pharmacoeconomic analyses. *Pharmacoeconomics* 1999;**15**:339–55.

59. Powe NR, Griffiths RI. The clinical-economic trial. Promise, problems and challenges. *Control Clin Trials* 1995;**16**:377–94.

60. Fayers PM, Hand DJ. Generalisation from phase III clinical trials: survival, quality of life, and health economics. *Lancet* 1997;**350**:1025–7.

61. Carr-Hill RA. The evaluation of health care. *Soc Sci Med* 1985;**21**:367–75.

62. Neymark N, Kiebert W, Torfs K, Davies L, Fayers P, Hillner B, *et al*. Methodological and statistical issues of quality of life (QoL) and economic evaluation in cancer clinical trials: report of a workshop. *Eur J Cancer* 1998; **34**:1317–33.

63. Tonkin AM. Issues in extrapolating from clinical trials to clinical practice and outcomes. *Aust NZ J Med* 1998;**28**:574–8.

64. Neymark N, Rosti G. Patient management strategies and transplantation techniques in European stem cell transplantation centers offering breast cancer patients high-dose chemotherapy with peripheral blood stem cell support: a joint report from the EORTC and EBMT. *Haematologica* 2000;**85**:733–44.

65. Postma MJ, Leidl R, Downs AM, Rovira J, Tolley K, Gyldmark M, *et al*. Economic impact of the AIDS epidemic in the European Community: towards multinational scenarios on hospital care and costs. *AIDS* 1993;**7**:541–53.

66. Rhodes G, Wiley M, Tomas R, Casas M, Leidl R. Comparing EU hospital efficiency using diagnostic-related groups. *Eur J Public Health* 1997;**7** (Suppl 3):42–50.

67. Nuffield Institute for Health, University of Leeds, NHS Centre for Reviews and Dissemination. Hospital volume and health care outcomes, costs and patient access. *Eff Health Care Bull* 1996; **2**(8):1–16.

68. Cowing T, Holtmann A, Powers S. Hospital cost analysis: a survey and evaluation of recent studies. *Adv Health Econ Health Serv Res* 1983;**4**:257–303.

69. Organisation for Economic Cooperation and Development (OECD). *OECD health data* 2000. Paris: OECD; 2000.

70. Soderlund N, Lavis J, Broomberg J, Mills A. The costs of HIV prevention strategies in developing countries. *Bull World Health Organ* 1993; **71**:595–604.

71. Cowper PA, DeLong ER, Peterson ED, Lipscomb J, Muhlbaier LH, Jolis JG, *et al*. Geographic variation in resource use for coronary artery bypass surgery. *Med Care* 1997;**35**:320–33.

72. Furlong WJ, Feeney DH, Torrance GW, Barr RD. The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. *Ann Med* 2001;**33**:375–84.

73. Williams A, Cookson R. Equity in health. In Culyer AJ, Newhouse JP, editors. *Handbook of health Economics*. Vol. 1B. Amsterdam: Elsevier Science; 2000. pp. 1863–1907.

74. Ramsay CR, Grant AM, Wallace SA, Garthwaite PH, Monk AF, Russell IT. Statistical assessment of the learning curves of health technologies. *Health Technol Assess* 2001;**5**(12).

75. Spiro HM. Diagnostic laparoscopic cholecystectomy. *Lancet* 1992;**339**:167–8.

76. Bridgman SA. Increasing operative rates for dysfunctional uterine bleeding after endometrial ablation (letter). *Lancet* 1994;**344**:893.

77. Schulman KA, Glick H, Buxton M, Sculpher M, Backhouse M, Bell L, *et al*. The economic evaluation of the FIRST study: design of a prospective analysis alongside a multinational phase III clinical trial. *Control Clin Trials* 1996; **17**:304–15.

78. Jonsson B, Johannesson M, Kjekshus J, Olsson AG, Pedersen TR, Wedel H. Cost-effectiveness of cholesterol lowering. Results from the Scandinavian Simvastatin Survival Study. *Eur Heart J* 1996;**17**:1001–7.

79. Holmes DR, Califf RM, Van de Werf F, Berger PB, Bates ER, Simoons ML. Differences in countries' use of resources and clinical outcome for patients with cardiogenic shock after myocardial infarction: results from the GUSTO trial. *Lancet* 1997; **349**:75–8.

80. Lacey L, Youle M, Trueman P. A prospective evaluation of the cost-effectiveness of adding lamivudine to zidovudine-containing antiretroviral treatment regimens in HIV infection. *Pharmacoeconomics* 1999;**15**:39–53.

81. Stalhammar NO, Carlsson J, Peacock R, Muller Lissner S, Bigard MA, Porro GB, *et al*. Cost effectiveness of omeprazole and ranitidine in intermittent treatment of symptomatic gastro-oesophageal reflux disease. *Pharmacoeconomics* 1999;**16**:483–97.

82. Mapelli V, Graf vd Schulenburg JM, Laaserm U. Economic evaluation of Lenograstim (glycosylated rHuG-CSF) in the treatment of inflammatory breast cancer for Germany and Italy. *Pharmacoeconomics* 1994;**6**:27–35.

83. Hakkaart-van Roijen L, Verboom P, Redekop WK, Touw KR, Rutten FF. The cost effectiveness of tapered versus abrupt discontinuation of oral cyclosporin microemulsion for the treatment of psoriasis. *Pharmacoeconomics* 2001;**19**:599–608.

84. Grieve R. *Can we generalise from costs measured in one country to another?* Paper presented at the International Health Economic Association 2nd World Conference, Rotterdam, June 1999.

85. Bryan S, Brown J, Warren R. Mammography screening: an incremental cost effectiveness analysis of two view versus one view procedures in London. *J Epidemiol Commun Health* 1995;**49**:70–8.

86. Pedersen TR, Kjekshus J, Berg K. Cholesterol lowering and the use of healthcare resources. Results of the Scandinavian Simvastatin Survival Study. *Circulation* 1996;**93**:1796–802.

87. Garson A Jr, Allen HD, Gersony WM, Gillette PC, Hohn AR, Pinsky WW, *et al*. The cost of congenital heart disease in children and adults. A model for multicenter assessment of price and practice variation. *Arch Pediatr Adolesc Med* 1994; **148**:1039–45.

88. Kennedy WA, Laurier C, Malo JL, L'Archeveque J, Ghezzo H, Contandriopoulos AP. Cost-efficacy vs cost-effectiveness: how does being chosen as a clinical trial subject affect the ability to generalise use and cost of health services to "real life" subjects? *Med Decis Making* 1999;**19**:531.

89. Leese B, Hutton J, Maynard A. A comparison of the costs and benefits of recombinant human erythropoietin (epoetin) in the treatment of chronic renal failure in 5 European countries. *Pharmacoeconomics* 1992;**1**:346–56.

90. Jansen RB, Capri S, Nuijten MJC, Burrell A, Marini MG, Hardens M. Economic evaluation of meloxicam (7.5 mg) versus sustained release diclofenac (100 mg) treatment for osteoarthritis: a cross-national assessment for France, Italy and the UK. *Br J Med Econ* 1997;**11**:1–2.

91. Arikian SR, Einarson TR, Kobelt Nguyen G, Schubert F. A multinational pharmacoeconomic analysis of oral therapies for onychomycosis. The Onychomycosis Study Group. *Br J Dermatol* 1994; **130**:35–44.

92. Einarson TR, Oh PI, Gupta AK, Shear NH. Multinational pharmacoeconomic analysis of topical and oral therapies for onychomycosis. *J Dermatolog Treat* 1997;**8**:229–35.

93. Jansen R, Redekop WK, Rutten FFH. Cost effectiveness of continuous terbinafine compared with intermittent itraconazole in the treatment of dermatophyte toenail onychomycosis. *Pharmacoeconomics* 2001;**19**:401–10.

94. Shear NH, Einarson TR, Arikian SR, Doyle JJ, Van Assche D. Pharmacoeconomic analysis of topical treatments for tinea infections. *Int J Dermatol* 1998;**37**:64–71.

95. Annemans L, Giaccone G, Vergnenegre A. The cost-effectiveness of paclitaxel (Taxol) + cisplatin is similar to that of teniposide + cisplatin in advanced non-small cell lung cancer: a multicountry analysis. *Anticancer Drugs* 1999; **10**:605–15.

96. Van Ineveld BM, Van Oortmarssen GJ, De Koning HJ, Boer R, Van der Maas PJ. How cost-effective is breast cancer screening in different EC countries? *Eur J Cancer* 1993;**29A**:1663–8.

97. Berger K, Fischer T, Szucs TD. Cost-effectiveness analysis of paclitaxel and cisplatin versus cyclophosphamide and cisplatin as first-line therapy in advanced ovarian cancer. A European perspective. *Eur J Cancer* 1998;**34**:1894–1901.

98. De Jonghe E, Murray CJL, Chum HJ, Nyangulu DS, Salomao A, Styblo K. Cost-effectiveness of chemotherapy for sputum smear-positive pulmonary tuberculosis in Malawi, Mozambique and Tanzania. *Int J Health Plann Manage* 1994;**9**:151–81.

99. Lorenzoni R, Pagano D, Mazzotta G, Rosen SD, Fattore G, De Caterina R, *et al*. Pitfalls in the economic evaluation of thrombolysis in myocardial infarction. The impact of national differences in the cost of thrombolytics and of differences in the efficacy across patient subgroups. *Eur Heart J* 1998;**19**:1518–24.

100. Caro J, Klittich W, McGuire A, Ford I, Pettitt D, Norrie J, *et al*. International economic analysis of primary prevention of cardiovascular disease with pravastatin in WOSCOPS. *Eur Heart J* 1999; **20**:263–8.

101. Grover SA, Coupal L, Zowall H, Alexander CM, Weiss TW, Gomes DRJ. How cost-effective is the treatment of dyslipidemia in patients with diabetes but without cardiovascular disease? *Diabetes Care* 2001;**24**:45–50.

102. Ghatnekar O, Persson U, Willis M, Odegaard K. Cost effectiveness of Becaplermin in the treatment of diabetic foot ulcers in four European countries. *Pharmacoeconomics* 2001;**19**:767–78.

103. Casciano J, Doyle J, Arikian S, Casciano R. The health economic impact of antidepressant usage from a payer's perspective: A multinational study. *Int J Clin Pract* 2001;**55**:292–9.

104. Simpson K, Hatzaindreu E, Anderson F. Cost effectiveness of antiviral treatment with Zalcitabine plus Zidovudine for AIDS patients with CD4+ counts less than 300/μL in 5 European countries. *Pharmacoeconomics* 1994;**6**:553–62.

105. Pinto CG, Lafuma A, Fagnani F, Nuijten MJ, Berdeaux G. Cost effectiveness of emedastine versus levocabastine in the treatment of allergic

conjunctivitis in 7 European countries. *Pharmacoeconomics* 2001;**19**:255–65.

106. Ament A, Baltussen R, Duru G, Rigaud-Bully C, De Graeve D, Ortqvist A, *et al*. Cost-effectiveness of pneumococcal vaccination of older people. A study in 5 western European countries. *Clin Infect Dis* 2000;**31**:444–50.

107. Neumann P. *How cost-effectiveness estimates change over time. Some examples and their implications for reimbursement policies*. Paper presented at the International Health Economics Association, 3rd World Conference, York. June 2001.

108. Remák E, Hutton J, Jones M, Zagari M. Changes in cost-effectiveness over time: the case of Epoetin Alfa for renal replacement therapy patients in the UK. *Eur J Health Econ* 2003;**4**:115–21.

109. Drummond MF, Davies L. Economic analysis alongside clinical trials. Revisiting the methodological issues. *Int J Technol Assess Health Care* 1991;**7**:561–73.

110. Rutten F. Economic evaluation and health care decision-making. *Health Policy* 1996;**36**:215–29.

111. Glick H. Strategies for economic assessment during the development of new drugs. *Drug Inf J* 1995;**29**:1391–1403.

112. Dixon S, Deverill M, Gannon M, Brazier J, Haggard M. Improving the validity of economic evaluations alongside controlled trials. *J Health Serv Res Policy* 1999;**4**:161–3.

113. Glasser JH, Chrzanowski RS. Medical technology assessment: adequate questions, appropriate methods, valuable answers. *Health Policy* 1988; **9**:267–76.

114. Baltussen R, Leidl R, Ament A. Real world designs in economic evaluation: Bridging the gap between clinical research and policy-making. *Pharmacoeconomics* 1999;**16**:449–58.

115. Haycox A, Dubois D, Butterworth M. Customising an international disease management model to the needs of individual countries: application to upper gastrointestinal disease. *Pharmacoeconomics* 1998; **2**:39–56.

116. Menzin J, Oster G, Davies L, Drummond MF, Greiner W, Lucioni C, *et al*. A multinational economic evaluation of rhDNase in the treatment of cystic fibrosis. *Int J Technol Assess Health Care* 1996;**12**:52–61.

117. Adams ME, McCall NT, Gray ET, Orza MJ, Chalmers TC. Economic analysis in randomized control trials. *Med Care* 1992;**30**:231–43.

118. Drummond MF. Experimental versus observational data in the economic evaluation of pharmaceuticals. *Med Decis Making* 1998;**18** (Suppl 2):S12–18.

119. Raikou M, Briggs A, Gray A, McGuire A. Centre-specific or average unit costs in multi-centre studies. Some theory and simulation. *Health Econ* 2000;**9**:191–8.

120. Schulman K, Burke J, Drummond M, Davies L, Carlsson P, Gruger J, *et al*. Resource costing for multinational neurologic clinical trials: methods and results. *Health Econ* 1998;**7**:629–39.

121. Glick HA, Orzol SM, Tooley JF, Polsky D, Mauskopf J. Design and analysis of unit cost estimation studies: How many types of hospitaliazation. How many countries? *Value Health* 2001;**4**:410.

122. Ludbrook A, Wordsworth S. *Making the results comparable in economic evaluations across countries: is purchasing power parity enough?* Paper presented at the International Health Economics Association, 3rd World Conference, York, June 2004.

123. Gray A. Are the results of cost-effectiveness studies internationally transferable? Paper presented to the Health Economists' Study Group, University of Aberdeen, 1991.

124. Sullivan SD, Liljas B, Buxton M, Lamm CJ, O'Byrne P, Tan WC, *et al*. Design and analytic considerations in determining the cost-effectiveness of early intervention in asthma from a multinational clinical trial. *Control Clin Trials* 2001;**22**:420–37.

125. Johnston K, Gerard K, Brown J. Generalizing costs from trials. Analyzing center selection bias in a breast screening trial. *Int J Technol Assess Health Care* 1998;**14**:494–504.

126. Mugford M, Hutton G, Fox Rushby J. Methods for economic evaluation alongside a multicentre trial in developing countries: a case study from the WHO Antenatal Care Randomised Controlled Trial. *Paediatr Perin at Epidemiol* 1998;**12**:75–97.

127. Coyle D, Lee KM. The problem of protocol driven costs in pharmacoeconomic analysis. *Pharmacoeconomics* 1998;**14**:357–63.

128. Oster G, Brook GM, Menzin J. A randomized trial to assess effectiveness and cost in clinical practice: rationale and design of the cholesterol reduction intervention study (CRIS). *Control Clin Trials* 1995; **16**:3–16.

129. Tunis SL, Johnstone BM, Kinon BJ, Barber BL, Browne RA. Designing naturalistic prospective studies of economic and effectiveness outcomes associated with novel antipsychotic therapies. *Value Health* 2000;**3**:232–42.

130. Caro JJ, Huybrechts KF, De Backer G, De Bacquer D, Closon MC. Are the WOSCOPS clinical and economic findings generalizable to other populations? A case study for Belgium. *Acta Cardiol* 2000;**55**:239–46.

131. Cook JR, Drummond M, Glick H, Heyse JF. Assessing the appropriateness of combining economic data from multinational clinical trials. *Stat Med* 2003;**22**:1955–76.

132. Hutton G. *Why can't health care costs be generalised within and between countries?* Paper presented at the International Health Economics Association, 3rd World Conference, York, June 2004.

133. Rutten-van Molken M, Van Doorslaer EKA, Denise-Till M. Cost-effectiveness analysis of formoterol versus salmeterol in patients with asthma. *Pharmacoeconomics* 1998;**14**:671–84.

134. Rutten-van Molken M, Van Doorslaer E. Pooling data from multinational cost-effectiveness analyses: illustrations from a CEA of two long-acting bronchodilators in asthma? Paper presented at the International Health Economic Association 2nd World Conference, Rotterdam, June 1999.

135. Koopmanschap MA, Touw KCR, Rutten FFH. Analysis of costs and cost-effectiveness in multinational trials. *Health Policy* 2001;**58**:175–86.

136. Sorensen S, de Lissovoy G, Mathur S, Duttagupta S. Projecting economic results of a European trial to the United States: issues and methods (abstract). *Value Health* 2000;**5**:374.

137. Pang F. The application of multilevel modelling and cluster analysis to multinational economic evaluation data. *Value Health* 1999;**2**:138.

138. Coyle D, Drummond MF. Analyzing differences in the costs of treatment across centers within economic evaluations. *Int J Technol Assess Health Care* 2001;**17**:155–63.

139. Rittenhouse BE. Exorcising protocol-induced spirits. Making the clinical trial relevant for economics. *Med Decis Making* 1997;**17**:331–9.

140. Jefferson T, Mugford M, Gray A, Demicheli V. An exercise on the feasibility of carrying out secondary economic analyses. *Health Econ* 1996; **5**:155–65.

141. Kinosian B, Glick H. Clinical trials and external validity. *Value Health* 1998;**1**:181–5.

142. Shih Y-C, Kauf TL. Reconciling decision models with the real world. An application to anaemia of renal failure. *Pharmacoeconomics* 1999;**15**:481–93.

143. Manning R, Claxton K. *Experimental and econometric solutions to selection bias*. Paper presented to the Health Economics Study Group, York, 1997.

144. McClellan M, Newhouse JP. The marginal cost-effectiveness of a medical technology. A panel instrumental-variables approach. *J Econom* 1997; **77**:39–64.

145. McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *JAMA* 1994;**272**:859–66.

146. Jefferson T, Demicheli V. Economic evaluation of influenza vaccination and economic modelling – can results be pooled? *Pharmacoeconomics* 1996; **9** (Suppl 3):67–72.

147. Spath HM, Carrere M-O, Fervers B, Philip T. Analysis of the eligibility of published economic evaluations for transfer to a given health care system: methodological approach and application to the French health care system. *Health Policy* 1999;**49**:161–77.

148. Welte R. Transfer of economic evaluation results between countries. Paper presented at the International Health Economic Association 3nd World Conference, York, July 2001.

149. Davis CE. Generalising from clinical trials. *Control Clin Trials* 1994;**15**:11–14.

150. Fenwick E, Claxton K, Sculpher M, Briggs A. *Improving the efficiency and relevance of health technology assessment: the role of decision analytic modelling. Centre for Health Economics Discussion Paper 179*. York: Centre for Health Economics, University of York; 2000.

151. Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. *Health Econ* 1996;**5**:513–24.

152. Claxton K, Neumann PJ, Araki SS, Weinstein MC. The value of information: an application to a policy model of Alzheimer's disease. *Int J Technol Assess Health Care* 2001;**17**:38–55.

153. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ* 1999;**18**:342–64.

154. Peto R, Collins R, Gray R. Large-scale randomised evidence: large and simple trials and overviews of trials. *J Clin Epidemiol* 1995;**48**:23–40.

155. Sculpher MJ, Poole L, Cleland J, Drummond M, Armstrong PW, Horowitz JD, *et al*. Low doses vs high doses of the angiotensin converting-enzyme inhibitor lisinopril in chronic heart failure: a cost-effectiveness analysis based on the Assessment of Treatment with Lisinopril and Survival (ATLAS) study. *Eur J Heart Fail* 2000;**2**:447–54.

156. Hoch JS, Briggs AH, Willan A. Something old, something new, something borrowed, something BLUE: a framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Econ* 2002;**11**:415–30.

157. Evans CJ, Crawford B. Data collection methods in prospective economic evaluations: how accurate are the results? *Value Health* 2000;**3**:277–86.

158. Brennan A, Akehurst R. Modelling in health economic evaluation: what is its place? what is its value? *Pharmacoeconomics* 2000;**17**:445–59.

159. Oostenbrink JB, Rutten-van Molken MP. Markov modelling to convert trial-based cost-effectiveness information to other countries – the Tiotropium example (abstract). *Value Health* 2001; **4**:422.

160. Rovira J. Standardizing economic appraisal of health technology in the European Community. *Soc Sci Med* 1994;**38**:1675–8.

161. Cook JR, Drummond M, Glick H, Heyse JF. *Analysing economic data from multi-national clinical trials. Issues and recommendations*. Paper presented at Health Economics Study Group meeting, July 1997.

162. Drummond MF, Jefferson T. Guidelines for authors and peer reviewers of economic submissions to the *BMJ*. *BMJ* 1996;**313**:275–83.

163. Drummond MF, Richardson WS, O'Brien BJ, Levine M, Heyland D. Users' guides to the medical literature XIII. How to use an article on economic analysis of clinical practice A. Are the results of the study valid? *JAMA* 1997;**277**:1552–7.

164. NHS Centre for Reviews and Dissemination. *Improving access to cost-effectiveness information for health care decision-making: the NHS Economic Evaluation Database*. 2nd ed. York: NHS CRD, University of York; 2001.

165. Stata. *Stata Statistical Software, Release 7.0*. College Station, TX: Stata Corporation; 2000.

166. Hutchins SS, Rosenthal J, Eason P, Swint E, Guerrero H, Hadler S. Effectiveness and cost-effectiveness of linking the special supplemental program for women, infants, and children (WIC) and immunization activities. *J Public Health Policy* 1999;**20**:408–26.

167. Webb A, Cunningham D, Scarffe JH, Harper P, Norman A, Joffe JK, *et al*. Randomised trial comparing epirubicin, cisplatin, and fluorouracil versus fluorouracil, doxorubicin, and methotrexate in advanced esophagogastric cancer. *J Clin Oncol* 1997;**15**:261–7.

168. Burton LC, Steinwachs DM, German PS, Shapiro S, Brant LJ, Richards TM, *et al*. Preventive services for the elderly: would coverage affect utilisation and costs under Medicare? *Am J Public Health* 1995;**85**:387–91.

169. Drummond N, Abdall M, Buckingham JK, Beattie JAG, Lindsay T, Osman LM, *et al*. Integrated care for asthma – a clinical, social, and economic evaluation. *BMJ* 1994;**308**:559–64.

170. Suver J, Arikian SR, Doyle JJ, Sweeney SW, Hagan M. Use of anesthesia selection in controlling surgery costs in an HMO Hospital. *Clin Ther* 1995;**17**:561–71.

171. Morrell CJ, Walters SJ, Dixon S, Collins KA, Brereton LML, Peters J, *et al*. Cost effectiveness of community leg ulcer clinics: randomised controlled trial. *BMJ* 1998;**316**:1487–91.

172. Grines CL, Marsalese DL, Brodie B, Griffin J, Donohue B, Constantini CR, *et al*. Safety and cost-effectiveness of early discharge after primary angioplasty in low risk patients with acute myocardial infarction. *J Am Coll Cardiol* 1998; **31**:967–72.

173. Gafni A, Goeree R, Myhr TL, Hannah ME, Blackhouse G, Willan AR, *et al*. Induction of labour versus expectant management for prelabour rupture of the membranes at term: an economic evaluation. *Can Med Assoc J* 1997; **157**:1519–25.

174. Mol BWJ, Hajenius PJ, Engelsbel S, Ankum WM, Hemrika DJ, Van der Veen F, *et al*. Treatment of tubal pregnancy in The Netherlands: an economic comparison of systemic methotrexate administration and laparoscopic salpingostomy. *Am J Obstet Gynecol* 1999;**181**:945–51.

175. Bailey TC, Ritchie DJ, McMullin ST, Kahn M, Reichley RM, Casabar E, *et al*. A randomized, prospective evaluation of an interventional program to discontinue intravenous antibiotics at two tertiary care teaching institutions. *Pharmacotherapy* 1997;**17**:277–81.

176. Sanchez Garcia M, Cambronero Galache JA, Lopez Diaz J, Cerda Cerda E, Rubio Blasco J, Gomez Aguinaga MA, *et al*. Effectiveness and cost of selective decontamination of the digestive tract in critically ill intubated patients: a randomized, double-blind, placebo-controlled, multicenter trial. *Am J Respir Crit Care Med* 1998;**158**:908–16.

177. Guthrie E, Moorey J, Margison F, Barker H, Palmer S, McGrath G, *et al*. Cost-effectiveness of brief psychodynamic-interpersonal therapy in high utilizers of psychiatric services. *Arch Gen Psychiatry* 1999;**56**:519–26.

178. Gournay K, Brooking J. The community psychiatric nurse in primary care: an economic analysis. *J Adv Nurs* 1995;**22**:769–78.

179. Nord E, Wisloff F, Hjorth M, Westin J. Cost–utility analysis of melphalan plus prednisone with or without Interferon-alpha 2b in newly diagnosed multiple myeloma. Results from a randomised controlled trial. *Pharmacoeconomics* 1997;**12**:89–103.

180. Torrance G, Walker V, Grossman R, Mukherjee J, Vaughan D, La Forge J, *et al*. Economic evaluation of ciprofloxacin compared with usual antibacterial care for the treatment of acute exacerbations of chronic bronchitis in patients followed for 1 year. *Pharmacoeconomics* 1999;**16**:499–520.

181. Glick H, Willke R, Polsky D, Llana T, Alves WM, Kassell N, *et al*. Economic analysis of tirilazad mesylate for aneurysmal subarachnoid hemorrhage. *Int J Technol Assess Health Care* 1998; **14**:145–60.

182. Smith TJ, Hillner BE, Neighbors DM, McSorley PA, Le Chevalier T. Economic evaluation of a randomized clinical trial comparing vinorelbine, vinorelbine plus cisplatin, and vindesine plus cisplatin for non-small-cell lung cancer. *J Clin Oncol* 1995;**13**:2166–73.

183. Charnas R, Luthi AR, Ruch W. Once daily ceftriaxone plus amikacin vs three times daily ceftazidime plus amikacin for treatment of febrile neutropenic children with cancer. *Pediatr Infect Dis J* 1997;**16**:346–53.

184. Bennett-Guerrero E, Sorohan JG, Gurevich ML, Kazanjian PE, Levy RR, Barbera AV, *et al*. Cost–benefit and efficacy of aprotinin compared with epsilon-aminocaproic acid in patients having repeated cardiac operations:a randomized, blinded clinical trial. *Anesthesiology* 1997;**87**:1373–80.

185. Serruys PW, van Hout B, Bonnier H, Legrand V, Eulogio Garcia, Macaya C, *et al*. Randomised comparison of implantation of heparin-coated stents with balloon angioplasty in selected patients with coronary artery disease (Benestent II). *Lancet* 1998;**352**:673–81.

186. Weaver WD, Reisman MA, Griffin JJ, Buller CE, Leimgruber PP, Henry T, *et al*. Optimum percutaneous transluminal coronary angioplasty compared with routine stent strategy trial (OPUS-1): a randomised trial. *Lancet* 2000; **355**:2199–203.

187. Mark DB HM, Califf RM, Naylor CD, Lee KL, Armstrong PW, Barbash G, *et al*. Cost-effectiveness of thrombolytic therapy with tissue plasminogen activator as compared with streptokinase for acute myocardial infarction. *N Engl J Med* 1995; **322**:1418–24.

188. Davey P, Craig AM, Hau C, Malek M. Cost-effectiveness of prophylactic nasal mupirocin in patients undergoing peritoneal dialysis based on a randomized, placebo-controlled trial. *J Antimicrob Chemother* 1999;**43**:105–12.

189. Scott WC, Cooper BC, Scott HM. Pharmacoeconomics evaluation of roxithromycin versus amoxycillin/clavulanic acid in a community-acquired lower respiratory tract infection study. *Infection* 1995;**23** (Suppl 1):S21–4.

190. Rajan R, Gafni A, Levine M, Hirsh J, Gent M. Very low-dose warfarin prophylaxis to prevent thromboembolism in women with metastatic breast cancer receiving chemotherapy: an economic evaluation. *J Clin Oncol* 1995;**13**:42–6.

191. Armstrong SH, Ruckley CV. Use of a fibrous dressing in exuding leg ulcers. *J Wound Care* 1997; **6**:322–4.

192. Hull RD, Raskob GE, Rosenbloom D, Pineo GF, Lerner RG, Gafni A, *et al*. Treatment of proximal vein thrombosis with subcutaneous low-molecular-weight heparin vs intravenous heparin. *Arch Intern Med* 1997;**157**:289–94.

193. Jonsson B, Cook JR, Pedersen TR. The cost-effectiveness of lipid lowering in patients with diabetes: results from the 4S trial. *Diabetologia* 1999;**42**:1293–301.

194. Liljas B, Stahl E, Pauwels RA. Cost effectiveness analysis of a dry powder inhaler (Turbuhaler) versus a pressurised metered dose inhaler in patients with asthma. *Pharmacoeconomics* 1997; **12**:267–77.

195. Mark DB, Harrington RA, Lincoff AM, Califf RM, Nelson CL, Tsiatis AA, *et al*. Cost-effectiveness of platelet glycoprotein IIb/IIIa inhibition with eptifibatide in patients with non-ST-elevation acute coronary syndromes. *Circulation* 2000;**101**:366–71.

196. Wootton R, Bloomer SE, Corbett R, Eedy DJ, Hicks N, Lotery HE, *et al*. Multicentre randomised control trial comparing real time teledermatology with conventional outpatient dermatological care: societal cost–benefit analysis. *BMJ* 2000;**320**:1252–6.

197. Ellis SL, Carter BL, Malone DC, Billups SJ, Okano GJ, Valuck RJ, *et al*. Clinical and economic impact of ambulatory care clinical pharmacists in management of dyslipidemia in older adults: the IMPROVE study. *Pharmacotherapy* 2000; **20**:1508–16.

198. Schulman KA, Burke J, Drummond M, Davies L, Carlsson P, Gruger J, *et al*. Resource costing for multinational neurologic clinical trials: methods and results. *Health Econ* 1998;**7**:629–38.

199. Walker D, Fox-Rushby J. Economic evaluation of parasitic diseases: a critique of the internal and external validity of published studies. *Trop Med Int Health* 2000;**5**:237–49.

200. Goldstein H, Rasbash J, Yang M, Woodhouse G, Pan H, Nuttal D, *et al*. A multi level analysis of school examination results. *Oxford Rev Educ* 1993; **19**:425–33.

201. DiPetre TA, Forristal JD. Multilevel models: methods and substance. *Annu Rev Sociol* 1994; **20**:331–57.

202. Kind P. The EuroQoL instrument: an index of health-related quality of life. In Spilker B, editor. *Quality of life and pharmacoeconomics in clinical trials*. Philadelphia, PA: Lippincott-Raven; 1996. pp. 191–201

203. Chaudahry MA, Sterns SC. Estimating confidence intervals for cost-effectiveness ratios: an example from a randomised trial. *Stat Med* 1996; **15**:1447–58.

204. Willan A, O'Brien B. Confidence intervals for cost-effectiveness ratios: an application of Fiellers' theorem. *Health Econ* 1996;**5**:297–305.

205. Briggs AH, Mooney CZ, Wonderling DE. Constructing confidence intervals for cost-effectiveness ratios: an evaluation of parametric and non-parametric techniques using Monte Carlo simulation. *Stat Med* 1999;**18**:3245–62.

206. Van Hout BA, Al MJ, Gordon GS, Rutten FFH. Costs, effects and c/e-ratios alongside a clinical trial. *Health Econ* 1994;**3**:309–19.

207. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Econ* 2001;**10**:779–89.

208. Briggs AH. Handling uncertainty in economic evaluation and presenting the results. In Drummond MF, McGuire A, editors. *Economic evaluation in health care. Merging theory with practice*. Oxford: Oxford University Press; 2001. pp. 172–214.

209. Stinnett A, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med Decis Making* 1998; **18** (Suppl 2):S68–80.

210. Briggs AH. A Bayesian approach to stochastic cost-effectiveness analysis. *Health Econ* 1999;**8**:257–62.

211. UK Prospective Diabetes Study Group. Cost effectiveness analysis of improved blood pressure control in hypertensive patients with type 2 diabetes: UKPDS 40. *BMJ* 1998;**317**:720–26.

212. Localio AR, Berlin JA, Ten Have TR, Kimmel SE. Adjustment for center in multicenter studies: an overview. *Ann Intern Med* 2001;**135**:112–23.

213. Snijders TAB, Bosker RJ. *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. London: Sage Publications; 1999.

214. Campbell MK, Mollison J, Grimshaw JM. Cluster trials in implementation research: estimation of intracluster correlation coefficients and sample size. *Stat Med* 2001;**20**:391–9.

215. Klar N, Donner A. Current and future challenges in the design and analysis of cluster randomised trials. *Stat Med* 2001;**20**:3729–40.

216. Aitkin M, Longford N. Statistical modelling issues in school effectiveness studies. *J Roy Stat Soc A* 1986;**149**:1–43.

217. Brown H, Prescott R. *Applied mixed models in medicine*. Chichester: John Wiley; 1998.

218. Cardoso AR. Wage differentials across firms: an application of multilevel modelling. *J Appl Econom* 2000;**15**:343–54.

219. Goldstein H, Healy M, Rasbash J. Multilevel time series model with applications to repeated measures data. *Stat Med* 1994;**13**:1643–55.

220. Di Prete TA, Grusky DB. *The multilevel analysis of trends with repeated cross-sectional data. Discussion Paper 88-7*. Chicago, IL: Economics Research Center/NORC; 1998.

221. Goldstein H, Yang M, Omar R, Turner R. Meta-analysis using multilevel models with an application to the study of class size effects. *Appl Stat* 2000;**49**:399–412.

222. Duncan C, Jones K, Moon G. Context, composition, and heterogeneity: using multilevel models in health research. *Soc Sci Med* 1998; **46**:97–117.

223. Rasbash J, Browne W, Goldstein H, Yang M, Plewis I, Healy M, *et al*. *A user's guide to MLwiN Version 2.1*. London: Institute of Education, University of London; 2000.

224. Garry R, Fountain J, Mason S, Napp V, Brown J, Hawe J. The eVALuate study: two parallel randomised trials, one comparing laparoscopic with abdominal hysterectomy, the other comparing laparoscopic with vaginal hysterectomy. *BMJ* 2004; **328**:129–35.

225. The Chartered Institute of Public Finance and Accountancy (CIPFA). *The Health Service Database 2000*. Croydon: CIPFA; 2000.

226. Joint Formulary Committee. *British National Formulary*. 38th ed. London: British Medical Association and the Royal Pharmaceutical Society of Great Britain; 1999.

227. Netten A, Dennett J, Knight J. *Unit Costs of Health and Social Care*. Canterbury: PSSRU; 2000.

228. Booth C. *Review Body for Nursing Staff, Midwives, Health Visitors and Professions Allied to Medicine 17th Report (Cm 4563)*. London: HMSO; 2000.

229. Gough CB. *Review Body on Doctors and Dentists Remuneration 29th Report (Cm 4562)*. London: HMSO; 2000.

230. Kind P, Hardman G, Macran S. *UK population norms for EQ-5D. Centre for Health Economics Discussion Paper 172*. York: Centre for Health Economics, University of York; 1999.

231. Garry R, Fountain J, Brown J, Manca A, Mason S, Sculpher M, *et al*. EVALUATE hysterectomy trial. A multicentre randomised trial comparing abdominal, vaginal and laparoscopic methods of hysterectomy. *Health Technol Assess* 2004;**8**(26).

232. The MRC Laparoscopic Groin Hernia Trial Group. Laparoscopic versus open repair of groin hernia: a randomised comparison. *Lancet* 1999; **354**:185–90.

233. The MRC Laparoscopic Groin Hernia Trial Group. Cost–utility analysis of open versus laparoscopic groin hernia repair: results from a multicentre randomised clinical trial. *Br J Surg* 2001;**88**:653–61.

234. Joint Formulary Committee. *British National Formulary*. 34th ed. London: British Medical Association and the Royal Pharmaceutical Society of Great Britain; 1997.

235. Netten A, Dennett J, Knight J. *Unit Costs of Health and Social Care*. Canterbury: PSSRU; 1997.

236. Packer M, Poole-Wilson PA, Armstrong PW, Cleland J, Horowitz JD, Massie BM, *et al*. Comparative effects of low and high doses of the angiotensin-converting enzyme inhibitor, lisinopril, on morbidity and mortality in chronic heart failure. *Circulation* 1999;**100**:2312–18.

237. Chartered Institute of Public Finance and Accountancy (CIPFA). *The Health Service Database 1999*. Croydon: CIPFA; 1999.

238. Joint Formulary Committee. *British National Formulary*. 36th ed. London: British Medical Association and the Royal Pharmaceutical Society of Great Britain; 1998.

239. Weinstein MC, Fineberg HV. *Clinical decision analysis*. Philadelphia, PA: Saunders; 1980.

240. Hunink M, Glaziou P, Siegel J, Weeks J, Pliskin J, Elstein A, *et al*. *Decision making in health and medicine. Integrating evidence and values*. Cambridge: Cambridge University Press; 2001.

241. Sculpher MJ, Torgerson D, Goeree R, O'Brien B. *A critical structured review of economic evaluations of interventions for the prevention and treatment of osteoporosis. Centre for Health Economics Discussion Paper No. 169*. York: Centre for Health Economics, University of York; 1999.

242. Royal College of Physicians of London. *Osteoporosis. Clinical guidelines for prevention and treatment*. London: Royal College of Physicians of London; 1999.

243. Office of Technology assessment (OTA), US Congress. *Effectiveness and costs of osteoporosis screening and hormone replacement therapy, volume I: cost effectiveness analysis (OTA-BP-H-160)*. Washington, DC: US Government Printing Office; 1995.

244. French F, Torgerson DJ, Porter RW. Overestimation of cost savings in prevention of osteoporosis (letter). *BMJ* 1994;**308**:138.

245. Ankjaer-Jensen A, Johnell O. Prevention of osteoporosis: cost-effectiveness of different pharmaceutical treatments. *Osteoporos Int* 1996; **6**:265–75.

246. Bombardier C, Maetzel A. Pharmacoeconomic evaluation of new treatments: efficacy versus effectiveness studies? *Ann Rheum Dis* 1999; **58**:182–5.

247. Geelhoed E, Harris A, Prince R. Cost-effectiveness analysis of hormone replacement therapy and lifestyle intervention for hip fracture. *Aust J Public Health* 1994;**18**:153–60.

248. Tosteson A, Rosenthal D, Melton L, Weinstein M. Cost effectiveness of screening perimenopausal white women for osteoporosis: bone densiometry and hormone replacement therapy. *Ann Intern Med* 1990;**113**:598–603.

249. Tosteson A, Weinstein M. Cost-effectiveness of hormone replacement therapy after the menopause. *Baillière's Clin Obstet Gynaecol* 1991; **5**:943–59.

250. Francis R, Anderson F, Torgerson D. A comparison of the effectiveness and cost of treatment for vertebral fractures in women. *Br J Rheumatol* 1995; **34**:1167–71.

251. Torgerson D, Reid D. Osteoporosis prevention through screening: will it be cost effective? *Baillière's Clin Rheum* 1993;**7**:603–22.

252. Torgerson D, Kanis J. Cost effectiveness of preventing hip fractures in the elderly population using vitamin D and calcium. *Q J Med* 1995; **88**:135–9.

253. Visentin A, Ciravegna R, Fabris F. The cost per avoided hip fracture by osteoporosis treatment in Italy. *Maturitas* 1997;**26**:185–92.

254. Cheung AP, Wren BG. A cost-effectiveness analysis of hormone replacement therapy in the menopause. *Med J Aust* 1992;**156**:312–16.

255. Daly E, Roche M, Barlow D, Gray A, McPherson K, Vessey M. HRT: an analysis of benefits, risks and costs. *Br Med Bull* 1992;**48**:368–400.

256. Daly E, Vessey M, Barlow A, Gray K, McPherson K, Roche M. Hormone replacement therapy in a risk-benefit perspective. *Maturitas* 1996;**23**:247–59.

257. Garton M, Cooper C, Reid D. Perimenopausal bone density screening – will it help prevent osteoporosis? *Maturitas* 1997;**26**:35–43.

258. Weinstein M. Estrogen use in postmenopausal women – costs, risks and benefits. *N Engl J Med* 1980;**303**:308–16.

259. Weinstein M, Schiff I. Cost-effectiveness of hormone replacement therapy in the menopause. *Obstet Gynecol Surv* 1983;**38**:445–54.

260. Rosner A, Grinna D, Torrance G, Bradley C, Adachi J, Sebaldt R, *et al*. Cost effectiveness of multi-therapy treatment strategies in the prevention of vertebral fractures in postmenopausal women with osteoporosis. *Pharmacoeconomics* 1998;**14**:559–73.

261. Coyle D, Cranney A, Lee KM, Welch V, Tugwell P. Cost-effectiveness of nasal calcitonin in postmenopausal women. Using Cochrane collaboration methods for meta analysis in economic evaluation. *Pharmacoeconomics* 2001; **19**:565–75.

262. Weinstein M, Tosteson A. Cost-effectiveness of hormone replacement. *Ann NY Acad Sci* 1990; **592**:162–73.

263. Palmer S, Sculpher M, Philips Z, Robinson M, Ginnelly L, Bakhai A, *et al*. *A cost-effectiveness model comparing alternative management strategies for the use of glycoprotein IIb/IIIa antagonists in non-ST-elevation acute coronary syndrome. Report to the National Institute for Clinical Excellence*. 2002. URL: http://www.nice.org.uk/Docref.asp?d=32030 (cited February 2003).

264. Robinson M, Ginnelly L, Sculpher MJ, Jones L, Riemsma R, Palmer S, *et al*. A systematic review

update of the clinical effectiveness and cost effectiveness of glycoprotein IIb/IIIa antagonists. *Health Technol Assess* 2002;**6**(25).

265. Boersma R, Harrington RA, Moliterno DJ, White H, Théroux P, Van de Werf F, *et al*. Platelet glycoprotein IIb/IIIa inhibitors in acute coronary syndromes: a meta-analysis of all major randomised clinical trials. *Lancet* 2002;**359**:189–98.

266. Department of Health. *Policy Appraisal and Health*. London: Department of Health; 1995.

267. British Cardiovascular Intervention Society. *BCIS audit returns of interventional procedures 2000*. London: BCIS; 2001. URL: http://www.bcis.org.uk/audit/oct01.html (cited February 2002).

268. Fox KAA, Poole-Wilson PA, Henderson RA, Clayton TC, Chamberlain DA, Shaw TR, *et al*. Interventional versus conservative treatment for patients with unstable angina or non-ST-elevation myocardial infarction: the British Heart Foundation RITA 3 randomised trial. *Lancet* 2002; **360**:743–51.

269. Collinson J, Flather MD, Fox KA, Findlay I, Rodrigues E, Dooley P, *et al*. Clinical outcomes, risk stratification and practice patterns of unstable angina and myocardial infarction without ST elevation: Prospective Registry of Acute Ischaemic Syndromes in the UK (PRAIS-UK). *Eur Heart J* 2000;**21**:1450–7.

270. Berkey CS, Hoaglin DC, Mosteller F, Colditz GA. A random effects regression model for meta-analysis. *Stat Med* 1995;**14**:395–411.

271. Sharp SJ, Thompson SG. Analysing the relationship between treatment effect and underlying risk in meta-analysis: comparison and development of approaches. *Stat Med* 2000; **19**:3251–74.

272. Spiegelhalter DJ, Thomas A, Best NG. *WinBUGS Version 1.2 User Manual*. Cambridge: MRC Biostatistics Unit; 2000.

273. Yusuf S, Zhao F, Mehta SR, Chrolavicius S, Tognoni G, Fox KK. Effects of clopidogrel in addition to aspirin in patients with acute coronary syndromes without ST-segment elevation. *N Engl J Med* 2001;**345**:494–502.

274. *ACC/AHA guideline update for the management of patients with unstable angina and non–ST-segment Elevation Myocardial Infarction*. 2002. URL: http://www.acc.org/clinical/guidelines/unstable/update_explantext.htm (cited 8 April 2002).

275. Gray D, Keating NA, Murdock J, Skene AM, Hampton JR. Impact of hospital thrombolysis policy on out-of-hospital response to suspected myocardial infarction. *Lancet* 1993;**341**:654–7.

276. Fryback DG, Chinnis JO, Ulviva JW. Bayesian cost-effectiveness analysis: an example using the GUSTO trial. *Int J Technol Assess Health Care* 2001; **17**:83–97.

277. Johannesson M, Weinstein S. On the decision rules of cost-effectiveness analysis. *J Health Econ* 1993; **12**:459–67.

278. Simon GE, Von Korff M, Heiligenstein JH, Revicki DA, Grothaus L, Katon W, *et al*. Initial antidepressant choice in primary care. Effectiveness and cost of fluoxetine vs tricyclic antidepressants. *JAMA* 1996; **275**:1897–902.

279. Johannesson M, O'Conor RM. Cost–utility analysis from a societal perspective. *Health Policy* 1997; **39**:241–53.

280. Sculpher M, Fenwick E, Claxton K. Assessing quality in decision analytic cost-effectiveness models: a suggested framework and example of application. *Pharmacoeconomics* 2000;**17**:461–77.

281. International Society for Pharmacoeconomics and Outcomes Research. *Principles for good practice in decision analytic modelling in health care evaluation. A statement from the ISPOR Taskforce on Modelling*. 2001. URL: http://www.ispor.org/workpaper/healthscience/TFModeling.pdf (cited February 2002).

282. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song TA. *Methods for meta-analysis in medical research*. Chichester: John Wiley; 2000.

283. Eddy DM. The confidence profile method: a Bayesian method for assessing health technologies. *Oper Res* 1989;**37**:210–28.

284. Ades AE, Cliffe S. Markov chain Monte Carlo estimation of a multi-parameter decision model: consistency of evidence and the accurate assessment of uncertainty. *Med Decis Making* 2002; **22**:359–71.

285. O'Brien BJ, Goeree R, Mohamed AH, Hunt R. *Economic evaluation of pharmaceuticals: Frankenstein's monster or vampire of trials?* Hamilton: CHEPA; 1995.

286. International Society for Pharmacoeconomics and Outcomes Research. *A report of the ISPOR Health Science Committee – Task Force on Good Research Practices – Modeling Studies*. 2000. URL: http://www.ispor.org/workpaper/healthscience/TFModeling.pdf

Appendix 1

Search strategies employed for electronic databases relating to the reviews detailed in Chapters 2–4

MEDLINE

Original search: 16 August 2000	SilverPlatter on ARC: 1980–August 2000
Update search one: 16 January 2001	SilverPlatter on ARC: September 2000–December 2000
Update search two: 5 February 2002	SilverPlatter on ARC: January 2001–November 2001

1. explode "Costs-and-Cost-Analysis"/ all subheadings
2. economic value of life in mesh
3. explode "Economics-Dental"/ all subheadings
4. explode "Economics-Hospital"/ all subheadings
5. explode "Economics-Medical"/ all subheadings
6. "Economics-Nursing"/ all subheadings
7. "Economics-Pharmaceutical"/ all subheadings
8. "Economics"/ all subheadings
9. explode "Fees-and-Charges"/ all subheadings
10. explode "Budgets"/ all subheadings
11. (cost or costs or costed or costly or costing) in ti ab
12. (economic* or pharmacoeconomic* or price* or pricing) in ti ab
13. #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12
14. letter in pt
15. editorial in pt
16. historical article in pt
17. #14 or #15 or #16
18. #13 not #17
19. animal in tg
20. human in tg
21. #19 not (#19 and #20)
22. #18 not #21
23. generalisability in ti ab
24. generalizability in ti ab
25. generalisable in ti ab
26. generalizable in ti ab
27. generalising in ti ab
28. generalizing in ti ab
29. external validity in ti ab
30. transportable in ti ab
31. transportability in ti ab
32. portable in ti ab
33. portability in ti ab
34. generalised in ti ab
35. generalized in ti ab
36. transferability in ti ab
37. transferable in ti ab
38. transferrable in ti ab
39. standardisation in ti ab
40. standardization in ti ab
41. adapted in ti ab
42. extrapolation in ti ab
43. extrapolated in ti ab
44. (applicable or applicability) in ti ab
45. (valid or validity) in ti ab
46. modelling in ti ab
47. modeling in ti ab
48. analysis in ti ab
49. hierach* in ti ab
50. hierarch* in ti ab
51. differ* in ti ab
52. assessment* in ti ab
53. (variance or variation) in ti ab
54. cross-national* in ti ab
55. crossnational* in ti ab
56. multicountry in ti ab
57. multi-country in ti ab
58. multinational in ti ab
59. multi-national in ti ab
60. multi-centre in ti ab
61. multicentre in ti ab
62. multicenter in ti ab
63. multi-center in ti ab
64. multilevel in ti ab
65. multi-level in ti ab
66. location* in ti ab
67. (country or countries) in ti ab
68. setting* in ti ab
69. health care systems in ti ab
70. locally in ti ab
71. (regions or regional or regionally) in ti ab
72. (hospitals or institutions) in ti ab
73. results in ti ab
74. (district or districts) in ti ab
75. nation wide in ti ab
76. geographical area in ti ab
77. nationwide in ti ab
78. global* in ti ab
79. world* in ti ab
80. europe in ti ab

81. (socio cultural or sociocultural) in ti ab
82. (socioeconomic or socio economic) in ti ab
83. (population* or nations) in ti ab
84. (context or contexts) in ti ab
85. future in ti ab
86. new drug* in ti ab
87. new technolog* in ti ab
88. trends in ti ab
89. new development* in ti ab
90. over time in ti ab
91. (clinical trials or study or research or theory) in ti ab
92. (health care or patient care or practice) in ti ab
93. #91 near #92
94. #23 or #24 or #25 or #26 or #27 or #28 or #29
95. #30 or #31 or #32 or #33 or #34 or #35 or #36 or #37 or #38 or #39 or #40 or #41 or #42 or #43 or #44
96. #45 or #46 or #47 or #48 or #49 or #50 or #51 or #52 or #53
97. #54 or #55 or #56 or #57 or #58 or #59 or #60 or #61 or #62 or #63 or #64 or #65
98. #66 or #67 or #68 or #69 or #70 or #71 or #72 or #73 or #74 or #75 or #76 or #77 or #78 or #79 or #80 or #81 or #82 or #83 or #84
99. #85 or #86 or #87 or #88 or #89 or #90
100. #98 or #99 or #93
101. #95 near #100
102. #96 near #97
103. #94 or #101 or #102
104. #103 and #22
105. PY >= "1985"
106. #104 and (PY >= "1985")

PREMEDLINE

Original search: 7 February 2002
SilverPlatter on ARC: up to 4 February 2002

1. (cost or costs or costed or costly or costing)
2. (economic* or pharmacoeconomic* or price* or pricing)
3. #1 or #2
4. generalisability
5. generalizability
6. generalisable
7. generalizable
8. generalising
9. generalizing
10. external validity
11. transportable
12. transportability
13. portable
14. portability
15. generalised
16. generalized
17. transferability
18. transferable
19. transferrable
20. standardisation
21. standardization
22. adapted
23. extrapolation
24. extrapolated
25. applicable or applicability
26. valid or validity
27. modelling
28. modeling
29. analysis
30. hierach*
31. hierarch*
32. differ*
33. assessment*
34. variance or variation
35. cross-national*
36. crossnational*
37. multicountry
38. multi-country
39. multinational
40. multi-national
41. multi-centre
42. multicentre
43. multicenter
44. multi-center
45. multilevel
46. multi-level
47. location*
48. country or countries
49. setting*
50. health care systems
51. locally
52. regions or regional or regionally
53. hospitals or institutions
54. results
55. district or districts
56. nation wide
57. geographical area
58. nationwide
59. global*
60. world*
61. europe
62. socio cultural or sociocultural
63. socioeconomic or socio economic
64. population* or nations
65. context or contexts
66. future
67. new drug*
68. new technolog*
69. trends

70. new development*
71. over time
72. clinical trials or study or research or theory
73. health care or patient care or practice
74. #72 near #73
75. #4 or #5 or #6 or #7 or #8 or #9 or #10
76. #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25
77. #26 or #27 or #28 or #29 or #30 or #31 or #32 or #33 or #34
78. #35 or #36 or #37 or #38 or #39 or #40 or #41 or #42 or #43 or #44 or #45 or #46
79. #47 or #48 or #49 or #50 or #51 or #52 or #53 or #54 or #55 or #56 or #57 or #58 or #59 or #60 or #61 or #62 or #63 or #64 or #65
80. #66 or #67 or #68 or #69 or #70 or #71
81. #79 or #80 or #74
82. #76 near #81
83. #77 near #78
84. #75 or #82 or #83
85. #84 and #3

EMBASE

Original search: 16 August 2000	SilverPlatter on ARC: 1980–July 2000
Update search one: 16 January 2001	SilverPlatter on ARC: August 2000–December 2000
Update search two: 5 February 2002	SilverPlatter on ARC: January 2001–January 2002

1. explode "economic-evaluation"/ all subheadings
2. cost effect* in ti ab
3. cost benefit* in ti ab
4. economic evaluation* in ti ab
5. pharmacoeconomic* in ti ab
6. cost util* in ti ab
7. #1 or #2 or #3 or #4 or #5 or #6
8. generalisability in ti ab
9. generalizability in ti ab
10. generalisable in ti ab
11. generalizable in ti ab
12. generalising in ti ab
13. generalizing in ti ab
14. external validity in ti ab
15. transportable in ti ab
16. transportability in ti ab
17. portable in ti ab
18. portability in ti ab
19. generalised in ti ab
20. generalized in ti ab
21. transferability in ti ab
22. transferable in ti ab
23. transferrable in ti ab
24. standardisation in ti ab
25. standardization in ti ab
26. adapted in ti ab
27. extrapolation in ti ab
28. extrapolated in ti ab
29. (applicable or applicability) in ti ab
30. (valid or validity) in ti ab
31. modelling in ti ab
32. modeling in ti ab
33. analysis in ti ab
34. hierach* in ti ab
35. hierarch* in ti ab
36. differ* in ti ab
37. assessment* in ti ab
38. (variance or variation) in ti ab
39. cross-national* in ti ab
40. crossnational* in ti ab
41. multicountry in ti ab
42. multi-country in ti ab
43. multinational in ti ab
44. multi-national in ti ab
45. multi-centre in ti ab
46. multicentre in ti ab
47. multi-center in ti ab
48. multicenter in ti ab
49. multilevel in ti ab
50. multi-level in ti ab
51. location* in ti ab
52. (country or countries) in ti ab
53. setting* in ti ab
54. health care systems in ti ab
55. locally in ti ab
56. (regions or regional or regionally) in ti ab
57. (hospitals or institutions) in ti ab
58. results in ti ab
59. (district or districts) in ti ab
60. nation wide in ti ab
61. geographical area in ti ab
62. nationwide in ti ab
63. global* in ti ab
64. world* in ti ab
65. europe in ti ab
66. (socio cultural or sociocultural) in ti ab
67. (socio economic or socioeconomic) in ti ab
68. (population* or nations) in ti ab
69. (context or contexts) in ti ab
70. future in ti ab
71. new drug* in ti ab
72. new technolog* in ti ab
73. trends in ti ab
74. new development* in ti ab
75. over time in ti ab
76. (clinical trials or study or research or theory) in ti ab

77. (health care or patient care or practice) in ti ab
78. #76 near #77
79. #8 or #9 or #10 or #11 or #12 or #13 or #14
80. #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27 or #28
81. #39 or #40 or #41 or #42 or #43 or #44 or #45 or #46 or #47 or #48 or #49 or #50
82. #51 or #52 or #53 or #54 or #55 or #56 or #57 or #58 or #59 or #60 or #61 or #62 or #63 or #64 or #65 or #66 or #67 or #68 or #69
83. #70 or #71 or #72 or #73 or #74 or #75
84. #78 or #83 or #82
85. #30 or #31 or #32 or #33 or #34 or #35 or #36 or #37 or #38
86. #80 near #84
87. #85 near #81
88. #86 or #87 or #79
89. #7 and #88
90. #89 and (PY >= "1985"

EconLit

Original search: 16 August 2000
Update search one: 16 January 2001
Update search two: 5 February 2002

SilverPlatter on ARC: 1969–June 2000
SilverPlatter on ARC: July 2000–November 2000
SilverPlatter on ARC: December 2000–December 2001

1. generalisability in ti ab
2. generalizability in ti ab
3. generalisable in ti ab
4. generalizable in ti ab
5. generalising in ti ab
6. generalizing in ti ab
7. external validity in ti ab
8. transportable in ti ab
9. transportability in ti ab
10. portable in ti ab
11. portability in ti ab
12. generalised in ti ab
13. generalized in ti ab
14. transferability in ti ab
15. transferable in ti ab
16. transferrable in ti ab
17. standardisation in ti ab
18. standardization in ti ab
19. adapted in ti ab
20. extrapolation in ti ab
21. extrapolated in ti ab
22. (applicable or applicability) in ti ab
23. (valid or validity) in ti ab
24. modelling in ti ab
25. modeling in ti ab
26. analysis in ti ab
27. hierach* in ti ab
28. hierarch* in ti ab
29. differ* in ti ab
30. assessment* in ti ab
31. (variance or variation) in ti ab
32. cross-national* in ti ab
33. crossnational* in ti ab
34. multicountry in ti ab
35. multi-country in ti ab
36. multinational in ti ab
37. multi-national in ti ab
38. multi-centre in ti ab
39. multicentre in ti ab
40. multi-center in ti ab
41. multicenter in ti ab
42. multilevel in ti ab
43. multi-level in ti ab
44. location* in ti ab
45. (country or countries) in ti ab
46. setting* in ti ab
47. health care systems in ti ab
48. locally in ti ab
49. (regions or regional or regionally) in ti ab
50. (hospitals or institutions) in ti ab
51. results in ti ab
52. (district or districts) in ti ab
53. nation wide in ti ab
54. geographical area in ti ab
55. nationwide in ti ab
56. global* in ti ab
57. world* in ti ab
58. europe in ti ab
59. (socio cultural or sociocultural) in ti ab
60. (socio economic or socioeconomic) in ti ab
61. (population* or nations) in ti ab
62. (context or contexts) in ti ab
63. future in ti ab
64. new drug* in ti ab
65. new technolog* in ti ab
66. trends in ti ab
67. new development* in ti ab
68. over time in ti ab
69. (clinical trials or study or research or theory) in ti ab
70. (health care or patient care or practice) in ti ab
71. #1 or #2 or #3 or #4 or #5 or #6 or #7
72. #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22
73. #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30 or #31

74. #32 or #33 or #34 or #35 or #36 or #37 or #38 or #39 or #40 or #41 or #42 or #43 or #44 or #45 or #46 or #47 or #48 or #49 or #50 or #51 or #52 or #53 or #54 or #55 or #56 or #57 or #58 or #59 or #60 or #61 or #62
75. #63 or #64 or #65 or #66 or #67 or #68
76. #69 near #70

Health Economic Evaluation Database (HEED) – CD-ROM

Original search: 17 August 2000 — Issue: August 2000
Update search one: 17 January 2001 — Issue: January 2000
Update search two: 6 February 2002 — Issue: January 2002

The following terms were searched in all data fields:

generalisability or generalizability or generalisable or generalizable or
generalising or generalizing or external validity or generalised or
generalized or applicability

CRD/CHE (University of York) Catalogue – CAIRS Internal software

Original search: 17 August 2000
Update search one: 17 January 2001
Update search two: 5 February 2002

The following terms were searched in all data fields:

S generalisability or generalizability or generalisable or generalizable or
generalising or generalizing or external validity or generalised or
generalized or applicability

Working Papers Database – Internet

http://econwpa.wustl.edu:80/months/hew
http://ideas.uqam.ca/
http://netec.mcc.ac.uk/BibEc.html

Original search 16 August 2000
Update search one: 16 January 2001
Update search two: 6 February 2002

The following terms were searched in all data fields:

generalisability or generalizability or generalisable or generalizable or
generalising or generalizing or external validity or generalised or
generalized or applicability

NHS Economic Evaluations Database (NHS EED) – Administration version of the database at CRD using CAIRS software

Original search: 16 August 2000 — Issue: August 2000
Update search one: 16 January 2001 — Issue: January 2000
Update search two: 5 February 2002 — Issue: January 2002

The following terms were searched in all data fields:

S generalisability or generalizability or generalisable or generalizable or
generalising or generalizing or external validity or generalised or
generalized or applicability

Conference Papers Index (CPI) – Dialog dialup service

Original search: 16 August 2000 — 1973–August 2000
Update search one: 19 January 2001 — 1973–January 2001
Update search two: 5 February 2002 — 1973–February 2002

1. COST(W)BENEFIT(W)ANALYS?
2. COST(W)EFFECTIVENESS(W)ANALYS?
3. COST(W)UTILITY(W)ANALYS?
4. PHARMACOECONOMIC?
5. ECONOMIC(W)EVALUATION?
6. GENERALISABILITY
7. GENERALIZABILITY
8. GENERALISABLE
9. GENERALIZABLE
10. GENERALISING

11. GENERALIZING
12. EXTERNAL(W)VALIDITY
13. TRANSPORTABLE
14. TRANSPORTABILITY
15. PORTABLE
16. PORTABILITY
17. GENERALISED
18. GENERALIZED
19. TRANSFERABILITY
20. TRANSFERABLE
21. TRANSFERRABLE
22. STANDARDISATION
23. STANDARDIZATION
24. ADAPTED
25. EXTRAPOLATION
26. EXTRAPOLATED
27. APPLICABLE OR APPLICABILITY
28. VALID OR VALIDITY
29. MODELLING
30. MODELING
31. ANALYSIS
32. HIERACH?
33. HIERARCH?
34. VARIANCE OR VARIATION
35. S1:S5
36. S6:S12
37. S13:S34
38. S35(3W)S37
39. S35 AND S36
40. S38 OR S39

Appendix 2

Summary tables of papers included in the review of methodological literature detailed in Chapters 3 and 4

(A) Conceptual factors affecting generalisability

Study	Objective	Systematic review methods	Vehicle	NHS-specific	Factors	Dimension
Carr-Hill[61]	To review well-known problems of the classic evaluation model, in particular the difficulty of generalising findings of an RCT to real clinical situations and the concentration of CEA on the margins of the health sector where criteria and objectives can be debated or where there is a question of relative exposure and power	No	Trial	No	B, E	1
Glasser[113]	To describe the process of technology assessment as a sequence of steps and the practical requisites of the diversity of questions, the economic and study design considerations are addressed in a systems analysis model of input–process–outcome	No	Not specific	No	D, P, Q, W	1
Stason[36]	To point to several real-world considerations for the application of CEA and to special challenges of international comparisons	No	Not specific	No	E, H, V, W, X	1, 2
Drummond[109]	To discuss issues in trial design, collection of resource use data, collection of outcome data and interpretation and extrapolation of results	No	Trial	No	A, N	1
Bonsel[45]	To describe theoretical aspects and practical aspects surrounding economic evaluations alongside cancer clinical trials	No	Trial	No	A, F, G, P, Q, W	1
Drummond[32]	To review arguments for and against standardisation in economic evaluation and make recommendations for further work	No	Not specific	No	A, E, P, Q	1
Mason[54]	To examine a recently reported league table and to identify methodological deficiencies	No	Not specific	Yes	A, E, M, P	1, 2
Tolley[47]	To review several European HIV–AIDS cost studies and outline the differences in methods used and where better approaches to cost estimation can be adopted	No	Not specific	No	A, F, P, Q, X	1
Bennett[55]	To summarise key methodological and practical issues associated with Phase 3 clinical trials that include clinical and economic outcomes	No	Trial	No	A, Q, W	1, 2
Briggs[16]	To review types of uncertainty that exist in economic evaluation and argue that some forms of uncertainty are not amenable to statistical methods	No	Not specific	No	A, B, C, D, E, O	1
Drummond[31]	To examine, with particular reference to the treatment of acid-related diseases, the reasons why cost-effectiveness data may vary by setting, to review the published literature to assess the comparative cost-effectiveness of products across countries and to provide some recommendations about the conduct and interpretation of economic studies on the international level	No	Model	No	A, E, H, L, P, S	1

continued

Study	Objective	Systematic review methods	Vehicle	NHS-specific	Factors	Dimension
Baker[37]	To address several important issues in the measurement of medical resource utilisation or direct medical costs when incorporating such measurements into clinical trials	No	Trial	No	B, H, J, N, T, W, Z	I
Bennett[56]	To describe the clinical and economic analysis of Phase 3 trials of GM-GSF as adjunct therapy in conjunction with high-dose chemotherapy and autologous stem cell transplantation for haematological malignancies	No	Trial	No	A, E, Q, T	I
Glick[111]	To consider issues in the design and planning of an economic assessment, including factors that influence the generalisability of the study's findings	No	Trial	No	B, N	I
O'Brien[285]	To review seven threats to the validity of pharmacoeconomic studies based on RCT data for both retrospective and prospective study designs	No	Trial	No	B, D, J, N, W, Z	I, 2
Powe[59]	To describe reasons why economic data collection and analysis are being considered in clinical trials, identify and discuss various strengths and limitations of different methods for gathering economic trial data for decision-making on appropriate use of healthcare interventions	No	Trial	No	A, E, K, N, Q, W	I, 2
Baltussen[35]	To show that age can have an important impact on the cost-effectiveness of medical interventions and to demonstrate that decision-makers' expectations concerning the effects of population ageing will be seriously impaired if this relationship is ignored	No	Not specific	No	H	I, 2
Baltussen[38]	To discuss the 'real world' relevance of results from economic evaluation as an additional step towards making results more useful to decision-makers	No	Trial (but relevant for other research designs)	No	B, C, D, F, I, J, K, U, P, Q, W	I, 2
Drummond[41]	To discuss issues facing the science, practice and future of pharmacoeconomics	No	Not specific	No	A, E, H, L, P, S	I
Drummond[1]	To discuss why economic data may not be easily transferable between countries or locations	No	Not specific	No	A, C, E, H, L, P, Q, S, Z	I
O'Brien[50]	To discuss the threats to transferring cost-effectiveness data from one country to another	No	Not specific	No	A, E, H, L, O, Q, S, Z	I
Rittenhouse[43]	To question whether conventional efficacy measurements in the form of premarketing, RCTs are appropriate for answering questions of interest to economists and to ask whether alternatives are available	No	Trial	No	H, Y, Z	I
Rutten[110]	To discuss specific problems with the use of cost-effectiveness information in health policy such as the use of cost-effectiveness profiles based on prospective clinical trials and the need to extrapolate from such information	No	Not specific	No	A, E, Q	I

continued

Study	Objective	Systematic review methods	Vehicle	NHS-specific	Factors	Dimension
Fayers[60]	To discuss the extent to which treatment effects, QoL findings and costs that are observed in an RCT can be generalised	No	Trial	No	A, B, W	1
Haycox[57]	To assess implications of incorporating economic evaluation into the design, implementation and interpretation of clinical trials	No	Trial	No	A, E, J, Q, Y, Z	1, 2
Mason[27]	To discuss the topic of generalisability in the light of three questions: (i) is the study technically good; (ii) do the results apply to my local decision-making context; and (iii) will the results apply generally in different jurisdictions with different perspectives?	No	Not specific	No	A, E, P, S, V, Z	1
Bryan[30]	To discuss issues in the application of cost-effectiveness information to the local population for which the provider or purchaser of healthcare is responsible	No	Not specific	Yes	A, L, Z	1
Coyle[46]	To discuss issues and, where appropriate, to provide recommendations for future studies of economic evaluations alongside clinical trials	No	Trial	No	B, D, F, I, Q, R	1, 2
Neymark[62]	To report in the proceedings of a symposium concerned with methodological issues in economic evaluation	No	Trial	No	E, G, J, N, W	1, 2
Tonkin[63]	To discuss issues in extrapolating from clinical trials to clinical practice	No	Trial	No	B, D, E, G, P	1
Dixon[112]	To illustrate how an economic evaluation alongside TARGET has been designed in order to avoid the threats to external validity that an awareness of clinical, operational and economic aspects of a service can reveal	No	Trial	No	B, J, N, W	1
Goeree[53]	To develop a conceptual framework for selecting hospitals for unit cost estimates in national and international multicentre trials and to test the impact of alternative hospital selection on the cost results	No	Trial	No	A, C, N, P, U, W, X	1
Goodman[39]	To assess the range and quality of the evidence base on the cost-effectiveness of malaria prevention and treatment in sub-Saharan Africa	No	Not specific	No	A, E, F, G, H, L, N, Q, W	1, 2
Revicki[40]	To summarise the key advantages and disadvantages of effectiveness or 'real world' trials relative to RCTs for evaluating pharmacoeconomic outcomes and for decision-making	No	Trial	No	B, C, E, F, H, K, P, Q, W	1
Rizzo[58]	To discuss 10 methodological hurdles of pharmacoeconomic analyses which researchers must overcome to perform successful analyses	No	Not specific	No	A, D, E, F, P, W	1, 2
Greiner[51]	To describe the economically relevant parameters that differ according to place and country of origin and to discuss current practice in order to solve these problems	No	Not specific	No	A, C, D, L, K, V, Q, S	1
Hughes[44]	To consider the effect of differences in compliance rates between normal clinical settings and the controlled environment of clinical trials on the cost-effectiveness of different drug therapies	No	Not specific	No	F	1

continued

Study	Objective	Systematic review methods	Vehicle	NHS-specific	Factors	Dimension
Neymark[64]	To examine extent of variation in current approaches in Europe to peripheral blood stem cell transplantation in breast carcinoma caused by differences in determining factors including: the type of institution, the centre cumulative experience with performing stem cell transplantations and the budget mechanisms of hospitals	No	Not specific	No	E, P	I

1, Location; 2, time.
A, absolute/relative costs (prices); B, artificial study conditions; C, capacity utilisation; D, case mix; E, clinical practice variation; F, compliance; G, culture/attitudes; H, demography; I, disease interaction; J, opportunity cost; K, economies of scale; L, epidemiology; M, exchange rates; N, geographical setting; O, health state valuations; P, healthcare resources; Q, healthcare system; R, historical differences; S, incentives; T, industry-related bias; U, joint production; V, perspective; W, skills/experience; X, technological innovation; Y, timing of assessment; Z, treatment of comparators.

(B) Empirical papers estimating variability in results between location and across time

Study	Objective	Clinical area/ (technology)	Location	Outcome measure	Variation factor in *Table 4*	Methods	Vehicle	Conclusions	Other information	Dimension
Arikian[91]	To determine the most cost-effective treatment of fingernail and toenail onychomycosis	Onychomycosis (Griseofulvin, itraconazole, ketoconazole, terbinafine)	A, B, CA, CH, D, E, F, GR, NL, IT, P, SF, UK	Cost per DFD	A, E, N	The effectiveness data (clinical success rates, relapse rates, side-effect rates) were taken from an international meta-analysis, and resource utilisation data (drugs, drug administration, routine medical care, laboratory tests, adverse drug reaction management costs) were established from country-specific guidelines and expert opinion	Model	Relative C/E of terbinafine versus other griseofulvin, itraconazole and ketoconazole varied by country. Terbinafine is associated with the lowest cost per DFD for all countries in the treatment of both fingernail and toenail infections. A ratio of 1 was used to describe the C/E of terbinafine. For fingernail, griseofulvin varied from 3.3/DFD (A) to 1.4/DFD (GR), itraconazole varied from 2.6/DFD (CH) to 1.4/DFD (UK), ketoconazole varied from 3.3/DFD (CA/SF) to 1.6/DFD (B). For toenail, griseofulvin varied from 3/DFD (A) to 1.5/DFD (I), itraconazole varied from 2.5/DFD (CA) to 1.05/DFD (UK), ketoconazole varied from 2.8/DFD(SF) to 1.55/DFD (P)	Multinational nature of the model is addressed using input from expert dermatologists and economists in each country providing input on clinical and cost parameters. The model is flexible in that it is able to accommodate the different reimbursement systems worldwide	1

continued

Study	Objective	Clinical area/ (technology)	Location	Outcome measure	Variation factor in *Table 39*	Methods	Vehicle	Conclusions	Other information	Dimension
Van Ineveld[96]	To assess the cost-effectiveness of breast cancer screening	Breast cancer screening (Screening, no screening)	E, F, NL, UK	Cost per screen	A, E, N, P, Q, W	The effectiveness data (incidence, mortality) were based on clinical trials in Sweden. Resource utilisation data (diagnosis, treatment, care) were established by expert opinion. Resources were valued using country-specific costs and standardised to UK sterling using healthcare-specific PPPs	Model	Spain=£9700/LYG, France=£5800/LYG, UK=£1800/LYG, The Netherlands = £2120/LYG	Differences between countries might include (i) the distinction between a specialised and an additive organisation of screening, (ii) the baseline in stage-specific survival, (iii) prognosis and attendance rates, (iv) false-positive rates, single- or dual-screening policy, (v) training of radiographers, radiodiagnosts, pathologists, (vi) equipment	I
Drummond[31]	To assess the cost-effectiveness of misoprostol for the prophylaxis of gastric ulcers in patients on long-term NSAID use	Gastric ulcers (Misoprostol, placebo)	B, F, UK, US	Cost	A, E, F, J, N	The effectiveness data (ulcer rates) were taken from a clinical trial. Resource utilisation data (diagnostic work-up, hospitalisations, surgery, hospital stay) were established by expert opinion, epidemiological surveys and routine hospital statistics. Resources were valued using country-specific costs and standardised to US dollars	Model	Expected costs (savings) per patient for 3 months prophylaxis at high dose: Belgium = $32, France = $61, UK = $55, US = $22; and at low dose: Belgium = $5, France = $15, UK = $3, US = $40	In the USA where the acquisition cost of misoprostol is highest, the overall economic results are most favourable. This is probably because of the relatively higher cost of other resources in the US healthcare system, especially surgery	I

continued

Study	Objective	Clinical area/ (technology)	Location	Outcome measure	Variation factor in *Table 39*	Methods	Vehicle	Conclusions	Other information	Dimension
De Jonghe[98]	To assess the cost-effectiveness of chemotherapy for pulmonary sputum smear-positive tuberculosis.	Pulmonary sputum smear-positive tuberculosis (Chemotherapy)	MA, MO, TA	Cost per life-year saved (LYS)	A, E, F, G, N, P, W	The effectiveness data (cure rate, mortality) were taken from cohort studies from 3 programmes in Malawi, Mozambique and Tanzania. Resource utilisation data (programme management, laboratory tests, drugs, hospitalisation, ambulatory) were established from accounting records. Resources were valued using country-specific costs and standardised to US dollars	Model	For short course hospitalisation, Malawi = $1.7/LYS, Mozambique = $2.6/LYS, Tanzania = $2.1/LYS. For standard with hospitalisation, Malawi = $2.4/LYS, Mozambique = $3.4/LYS, Tanzania = $3.1/LYS. For ambulatory short course, Malawi = $1.1/LYS, Mozambique = $0.9/LYS, Tanzania = $1.1/LYS. For ambulatory standard, Malawi = $1.3/LYS, Mozambique = $0.9/LYS, Tanzania = $1.2/LYS	The costs per patient treated for different regimens are remarkably similar except for differences in the costs of food and labour. Generalising the results to other countries is questionable because of differences in national incomes per capita where the cost of labour is higher in dollar terms	I
Garson[87]	To assess the cost and practice variation for a chronic disease among six US geographically representative centres	CoHD Various	USA	Cost	A, E, N	Resource utilisation data (clinic visits, hospitalisations, medication) were taken from an observational study at 6 sites. Since treatment of CoHD involves more than one medical service, cost variation is addressed by generalising from three standard cost accounting relationships to accommodate multiple modalities (volumes), prices and sites. These were (i) price variance (= price for a service at a centre-mean price for that service for six centres) × frequency	Observa-tional	A 55% variation exists among six centres in the charges for congenital heart disease between birth and 21 years. Charges and (mortality) were for institution A = $63,283 (8%), B = $57,071 (8%), C = $59,187 (5%), D = $62,346 (4%), E = $47,515 (8%), F = £73,606 (7%)	Variation is unrelated to outcome (mortality) and approximately 50% can be accounted for by practice pattern. Greatest variation is in the use of complex clinic visits and cardiac catherisation. The study has shown that, by means of a multicentre approach, practice variation can be assessed and related to outcome	I

continued

Study	Objective	Clinical area/ (technology)	Location	Outcome measure	Variation factor in *Table 39*	Methods	Vehicle	Conclusions	Other information	Dimension
						for that service at that centre; (ii) practice pattern variance (= frequency of a service at a centre – mean frequency for that service for six centres) × mean price for that service; (iii) cost variance (= price variance + practice pattern variance)				
Bennett[56]	To assess the costs of GM-GSF as adjunct therapy in conjunction with high-dose chemotherapy and autologous stem cell transplantation for haematological malignancies	Cancer (GM-CSF)	F, US	Cost	A, B, E, N, Q, T, V	The effectiveness data (duration of neutropenia, time from reinfusion of bone marrow and/or peripheral blood stem cells to hospital discharge) and resource utilisation data (hospitalisations, radiographs, laboratory tests, blood products, hyperalimentation procedures) were taken from two clinical trials. Resource utilisation data (hospitalisations, radiographs, laboratory tests, blood products, hyperalimentation) were collected at two tertiary care centres in Paris and New York	Trial	Clinical results were similar in both Paris and New York, but economic results differed. GM-CSF was associated with a marked decrease in severe neutropenia duration. Significant savings in terms of fewer days in hospital, laboratory tests, radiographs for GM-CSF patients were noted for the New York hospital, but not identified at the Paris hospital. Mean costs of care were $61,815 for GM-CSF and $68,640 for placebo	A random effects model was used to account for institutional variation in overall clinical results: $y = az + bi + e$, where y is the outcome (days of neutropenia; length of hospital stay) for a subject in hospital i ($i = 1 - 6$), z is a treatment indicator (0 = placebo, 1 = GM-CSF), a is an unknown parameter, b is a random hospital effect and e is a random error term (mean = 0)	I

continued

Study	Objective	Clinical area/ (technology)	Location	Outcome measure	Variation factor in *Table 39*	Methods	Vehicle	Conclusions	Other information	Dimension
Schulman[34]	To assess the cost-effectiveness of epoprostenol (Flolan) versus usual care in the treatment of congestive heart failure	Congestive heart failure (Epoprostenol, usual care)	CA, US, 12 EU	Cost	A, B, N	The effectiveness data (mortality) and resource utilisation data (hospitalisations, procedures, inpatient/outpatient, nursing, laboratory tests, nursing home, hospice, spa) were taken from the Flolan International Randomised Survival Trial (FIRST). Resources were valued using US costs from a single hospital.	Trial	The authors did not report results by location although variability may have existed. Within the trial, costs and (survival) were epoprostenol = $16,819 (4.3 months), usual care = $11,797 (4.7 months). At 12 months, projected costs and survival were epoprostenol = $27,747 (5.8 months), usual care = $22,476 (7.1 months)	Unit costs from one centre may not reflect either the average unit costs among study centres or the true variation in unit costs. However, adjusted cost data may be useful to different audiences. European unit costs are thought to be lower than US costs and an increase in variance to reflect the effects of multiple sets of unit cost estimates	I
Menzin[116]	To assess the cost-effectiveness of recombinant human deoxyribonuclease (rhDNase) to improve pulmonary function in patients with cystic fibrosis	RTI (RhDNase)	D, F, IT, UK, US	Cost	A, E, N	The effectiveness data (change in pulmonary function (FEV1), incidence of RTIs requiring parenteral therapy) and resource utilisation data (rate of hospitalisation, antibiotics, length of stay) were taken from a US clinical trial. Further resource information was ascertained through expert opinion on 'practice pattern' parameters believed likely to vary across countries such as the likelihood of hospitalisation and the associated mean length of stay in hospital. Resources were valued using country-specific costs and standardised to US dollars. A decision model was used as a means of extrapolating from the US trial data to the other locations.	Model	Savings were demonstrated in all countries: France = $1070, Germany = $934, Italy = $864, UK = $690. After adjustments were made based on the variability in two key parameters, Italy = $600, France = $860. In Germany, there was a very small increase in savings from treating RTIs	It was recognised that important differences may exist between the USA and other countries. Therefore, data were collected in each country to ascertain whether practice patterns varied and estimates of the projected economic impact of rhDNase were adjusted accordingly	I

continued

Study	Objective	Clinical area/ (technology)	Location	Outcome measure	Variation factor in *Table 39*	Methods	Vehicle	Conclusions	Other information	Dimension
Einarson[92]	To assess the cost-effectiveness of topical lacquers and oral agents for the treatment of mild to moderate onychomycosis	Onychomycosis [Topical lacquers (amorolfine, ciclopirox), oral agents (griseofulvin, itraconazole, terbinafine)]	CA, D, E, F, IT, UK	Cost per regimen; total expected costs of therapy; cost per SFD	A, E, N	The effectiveness data (cure rates, failure rates, relapse rates) were taken from a meta-analysis of 33 studies comprising 58 treatment arms. Resource utilisation data (drugs, pharmacist, physician visits, laboratory tests, nail avulsion surgery) were established by expert opinion. Resources were valued using country-specific costs and standardised to US dollars	Model	Ciclopirox as first-line therapy had the lowest expected cost and lowest cost per SFD followed by amorolfine, terbinafine and itraconazole in all countries except the UK and Spain. Ciclopirox ranged from $0.20/SFD (CA) to $0.50/SFD (UK); amorolfine ranged from $0.18/SFD (IT) to $0.48/SFD (UK); itraconazole ranged from $0.25/SFD (CA) to $0.45/SFD (F); terbinafine ranged from $0.25/SFD (CA) to $0.39/SFD (UK); griseofulvin ranged from $0.60/SFD (F) to $1.50/SFD (D)	Variations were observed in the UK and Spain since costs were extrapolated from the price in Germany and this assumption is likely to be an overestimation. The cost of failure was much higher in the UK than for other countries	I
Jonsson[78]	To assess the cost-effectiveness of simvastatin 20–40 mg once daily versus usual care for the treatment of coronary heart disease.	Coronary heart disease (Simvastatin/ usual care)	AU, B, D, E, F, IT, NZ, NO, P, SW, UK, US	Cost per LYS	A, E, N, P	The effectiveness data (mortality) and resource utilisation data (hospitalisations, revascularisation procedures, medications) were taken from the Scandinavian Simvastatin Survival Study (4S) clinical trial. Resources were valued using country-specific costs and standardised to UK sterling.	Trial	C/E ratios were similar across all countries: Sweden = £5502/LYS, Norway = £6361/LYS, Belgium = £5165/LYS, France = £4137/LYS, Germany = £7827/LYS, Italy = £5869/LYS, Portugal = £8312/LYS, Spain = £6148/LYS, UK = £6983/LYS, Australia = £5970/LYS, New Zealand = £8824/LYS	Estimates based on 4S utilisation data and local unit costs revealed that C/E of simvastatin was constant in various countries despite international differences in hospitalisation and drug costs. Procedures can be highly variable according to location. PTCA in the 4S countries were similar to the European average, but the rates of CABG were slightly higher than in Europe	I

continued

Study	Objective	Clinical area/ (technology)	Location	Outcome measure	Variation factor in *Table 39*	Methods	Vehicle	Conclusions	Other information	Dimension
Holmes[79]	To compare use of resources for patients with cardiogenic shock in several countries and to assess the association between use of resources and clinical outcome	MI	USA + other countries	Resources	E, H, N, P	The effectiveness data and resource utilisation data (pulmonary artery, catheterisation, cardiac catheterisation, intravenous inotropic agents, ventilatory support, intra-aortic balloon counter-pulsation (IABP), PTCI, CABG) were taken from the GUSTO-I trial	Trial	Aggressive diagnostic and therapeutic procedures were used more aggressively in the USA than in other countries: cardiac catherisation (58 vs 23%), IABP (35 vs 7%), right-hand catherisation (57 vs 22%), ventilatory support (54 vs 38%), PTCA (26 vs 8%). Adjusted 30-day mortality was significantly lower among patients treated in the USA than elsewhere (50 vs 60%, $p < 0.001$). 1 year mortality was 56 vs 70%	The difference in outcome (mortality) may be due to greater use of invasive diagnostic and therapeutic interventions in the USA	I
Lacey[80]	To examine the cost-effectiveness of adding lamivudine to zidovudine containing antiretroviral regimens	HIV (Lamivudine/zidovudine; lamivudine/loviride/zidovudine; lamivudine/placebo/zidovudine; zidovudine/didanosine; zidovudine/zalcitabine)	AU, CA, EU, SA	Cost per DPA	A, B, E, N	The effectiveness data (HIV-related events, progressions to AIDS/death) and resource utilisation data (hospitalisations, outpatient visits, medications, adverse events) were taken from the CAESAR clinical trial. Resources were valued using UK and German unit costs	Trial	D = DM22,405/DPA, UK = £12,030/DPA. The additional costs per HIV-related event avoided over the duration of the trial were D = DM8869, UK = £4762	The rationale for pooling was that no statistically significant differences were found for hospitalisation rates. When country was controlled in the comparison of resource utilisation across treatment groups, treatment effects were maintained in the majority of cases, suggesting consistency in treatment differences across countries. Lengths of stay for a basket of HIV-related illness events were similar between the CAESAR trial and a German nationwide database	I

continued

Study	Objective	Clinical area/ (technology)	Location	Outcome measure	Variation factor in *Table 39*	Methods	Vehicle	Conclusions	Other information	Dimension
Grieve[84]	To compare the resource use, total costs and survival associated with providing stroke care in different European settings	Stroke	A, D, F, LI, PO, UK	Costs, resources	A, E, P	For each stroke patient hospitalised during 1996, resource use (hospital stay, investigations, medical time) were collected. Resources were valued using centre-specific unit costs and standardised to US dollars	Trial	The costs of stroke management varied across Europe. The mean length of stay ranged from 12 days in France to 34 days in the UK. The average nursing time per day ranged from 113 minutes in Poland to 318 minutes in Austria. Also unit costs were higher in western than in eastern Europe (e.g. a doctor cost $34.21 per hour in France compared with $1.32 per hour in Lithuania). The lowest mean cost per case in Lithuania ($880, 95% CI 730 to 1030) and the highest in Austria ($8336, 95% CI 6638 to 10033)	It can be concluded from this study that hospital costs measured in one European country do not seem generalisable to another and the pooled data from multinational economic evaluations may not be applicable to the local setting	I
Jansen[90]	To compare the cost consequences of meloxicam (7.5 mg) versus diclofenac SR (100 mg) for the treatment of osteoarthritis	Osteoarthritis (Meloxicam, diclofenac SR)	F, IT, UK	Costs	A, E, N	The effectiveness data (adverse event rates) were taken from clinical trials, and resource utilisation data were determined from a medical database (UK), previously published literature (UK, France) and expert opinion (UK, Italy, France). Resources were valued using country-specific costs	Model	Substantial differences in resource use exist between countries for the same event. Hospitalisation for ulcer treatment is: F = 13.2–30.8 days and UK = 8.6 days. Potential cost savings from the payer perspective were: F = $18.16 (32%), IT = $4.34 (5%), UK = 13.89 (24%) per patient per 30-day treatment	The large cost savings in France were due to elevated hospitalisation costs. The modest cost savings in Italy were believed to be due to the higher cost of OA patients without a GI event and the Diagnostic Related Groups (DRG)-based costing methodology	I

continued

Study	Objective	Clinical area/ (technology)	Location	Outcome measure	Variation factor in *Table 39*	Methods	Vehicle	Conclusions	Other information	Dimension
Annemans[95]	To assess the cost-effectiveness of paclitaxel + cisplatin (TAXCIS) versus teniposide + cisplatin (TENCIS)	Advanced non-small cell lung cancer Paclitaxel + cisplatin (TAXCIS); teniposide + cisplatin (TENCIS)	B, E, F, NL	Cost per extra responder	A, E, N	The effectiveness data (partial remission, complete remission) were taken from two databases (EORTC, BMS) based on the same multinational clinical trial. The databases had small differences depending on the interpretation of some outcomes. Resource utilisation data (drug, administration, haematological adverse events) were established by expert opinion and patient chart analysis. Resources were valued using country-specific costs and standardised to US dollars	Model	Clear differences exist in practice and costs between the four countries. There is more preventative use of G-CSF in The Netherlands and Belgium; the costs of testing are lower than in Spain or France and the unit costs and hospitalisation rates are higher in France. C/E ratios based on the EORTC were: The Netherlands = $33,323/responder, Belgium = $19,409/ responder; Spain = $20,559/responder, France = $14,202/responder; and based on BMS were: The Netherlands = $33,550/responder, Belgium = $17,775/ responder; Spain = $20,394/responder, France = $12,323/ responder	Medical practice variation influenced the C/E ratios, but not the conclusions. The study used the Delphi approach to collect data regarding medical practice	I
Kennedy[88]	To compare the use and cost of asthma-related health services in trial and non-trial patients	Asthma (ICSTs)	CA (RCT and non-RCT)	Costs	A, B	Resource utilisation data (ICSTs, emergency room, GP) were based on 75 patients enrolled in a clinical trial and 51 patients who did not participate in the trial. Comparisons were performed using logistic regression controlling for age group, asthma severity, year of data collection and geographic location (Montreal or Quebec City)	Trial/ observa-tional	Trial patients were more likely to use higher (400 μg or more) daily doses of ICSTs than non-trial patients (OR: 3.1, 85% CI 1.6 to 6.2). Trial patients were less likely to visit the emergency department than non-trial patients (OR 0.4, 85% CI 0.2 to 0.8) and less likely to have 2 or more GP visits per year (OR 0.3, 85% CI 0.2 to 0.6). Log transformed total asthma-related costs did not differ between trial and non-trial	Certain categories of services (including ICSTs, emergency department physician visits, GP visits) differed in individuals taking ICSTs for their asthma, whether or not they have been enrolled in a clinical trial, but could not conclude that there was a difference in the total cost of asthma-related health services.	I

continued

Study	Objective	Clinical area/ (technology)	Location	Outcome measure	Variation factor in *Table 39*	Methods	Vehicle	Conclusions	Other information	Dimension
Stalhammar[81]	To assess the cost-effectiveness of omeprazole and rantidine when used as initial therapy in an intermittent treatment strategy for management of patients with symptomatic GORD with or without erosive oesophagitis	Gastro-oesophageal reflux disease (Omeprazole, ranitidine)	D, E, F, IR, IT, UK	Cost	A, E, N	The effectiveness data (symptoms) and resource utilisation data (medication, examination, transportation, productivity loss) were taken from a multinational clinical trial. In two separate analyses, trial-wide and country-specific resources were valued using country-specific unit costs	Trial	Estimated costs were found to be lower for both dosages of omeprazole than for ranitidine in all countries except Germany. UK = £356 (omeprazole 20 mg), £376 (omeprazole 10 mg), £527 (ranitidine 150 mg bd); Ireland = £1577 (omeprazole 20 mg), £1592 (omeprazole 10 mg), £1521 (ranitidine 150 mg bd); Germany = DM1186 (omeprazole 20 mg), DM1388 (omeprazole 10 mg), DM1111 (ranitidine 150 mg bd); France = FF3277 (omeprazole 20 mg), FF3236 (omeprazole 10 mg), FF3500 (ranitidine 150 mg bd); Italy = L1,081,337 (omeprazole 20 mg), L1,044,697 (omeprazole 10 mg), L1,154,754 (ranitidine 150 mg bd); Spain = Pta87,557 (omeprazole 20 mg), Pta86,528 (omeprazole 10 mg), Pta80,007 (ranitidine 150 mg bd). Using country-specific data led to substantial changes in some of the countries: in France and Ireland, ranitidine became the least costly alternative and in Italy and Spain, omeprazole 20 mg became the least costly alternative	In a multinational study, important questions arise about the extent trial-wide data on costs can be assumed to apply to all countries. It may be possible to test for homogeneity among the countries, but there is the obvious risk of low power. Trial-wide data on effectiveness and quantities of medication, visits and endoscopies were assumed to apply to all countries, whilst the extent of sick leave and the mode and time for transportation were allowed to vary between the countries. Sensitivity analysis was performed where country-specific data on resources were used	I

continued

Study	Objective	Clinical area/ (technology)	Location	Outcome measure	Variation factor in *Table 39*	Methods	Vehicle	Conclusions	Other information	Dimension
Leese[89]	To assess the cost-effectiveness of human erythropoietin (epoetin) in the treatment of anaemia arising from chronic renal failure	Anaemia [Human erythropoietin (epoetin), usual care]	D, E, F, IT, UK	Cost per QALY	A, D, E, N, Q	The health outcome data (utilities) were generated prospectively in small samples of patients in Italy and Spain and reinforced by assessments of patients by medical staff in Germany, France and the UK. Resource utilisation data (epoetin, transfusions, side-effects) were established by expert opinion. Resources were valued using country-specific costs and standardised to US dollars. The same study was conducted by separate groups in each country using a common methodological approach	Model	Germany = $171,123/QALY, France = $124,924/QALY, Spain = 348,992/QALY, Italy = 119,712/QALY, UK = 176,075/QALY	Each country's different approach to selecting patients for epoetin treatment and different healthcare financing arrangements led to varying treatment regimens and costs. UK shows the largest average QALY gain, possibly reflecting the greater average severity of anaemia in the limited number of patients given epoetin in the UK	I
Mapelli[82]	To assess costs of lenograstin as an adjunct to chemotherapy for the treatment of women with inflammatory breast cancer	Inflammatory breast cancer (Lenograstin, placebo)	D, IT	Cost	A, N	The effectiveness data (incidence of complete response) and resource utilisation data (antibiotic therapy, chemotherapy, infections, adverse events) were taken from a trial involving 10 French hospitals. Resources were valued using German or Italian unit costs	Trial	Excess costs for lenograstim were: D = DM4166, IT = L2.4 million; and for placebo were: D = DM5960, IT = L3.7 million. Assuming costs of chemotherapy were the same for both groups, cost saving in the lenograstin group would be 30% in Germany and 34% in Italy	The underlying assumption is that patients treated in French centres would have been treated in a similar way in German and Italian hospitals. Caution should be used in extending the results of this study to other situations since it is difficult at this time to assess the cost-effectiveness of lenograstin in this indication since long-term results are awaited	I

continued

Study	Objective	Clinical area/ (technology)	Location	Outcome measure	Variation factor in *Table 39*	Methods	Vehicle	Conclusions	Other information	Dimension
Haycox[115]	To assess international transferability of a disease model for UGI disease to different countries	UGI disease [GORD, ulcer, motility disorders] [Proton-pump inhibitors (PPIs), H2 receptor angonists]	A, AU, D, IT, NL.SW, UK, US	Cost	A, E, M, N, P	The resource utilisation data (medications, GP, outpatient, procedures, inpatient, laboratory tests) were established by expert opinion. Resources were valued using country-specific costs and standardised to UK sterling	Model	Wide variability exists between countries in the cost of treating patients for UGI disease (35% GORD, 32.5% ulcer, 32.5% motility disorders): UK = £174.52, Sweden = £312.40, Switzerland = £283.04, Germany = £176.11 A dichotomy exists between the UK and Germany ('the cheap countries') and Sweden and Switzerland ('the expensive countries'). In all major cost categories, the cost of treating patients in the UK and Germany is significantly lower than that in Sweden and Switzerland	The model was sensitive to cost but comparatively insensitive to changes to transition probabilities underlying the model. Much variation can be attributed to differences in the availability and costs of resources between countries. Given the proven transferability of the model, it can be reasonably expected that significant cost advantages underlying implementation in the UK would be equally applicable to other European countries	I
Caro[100]	To assess the cost-effectiveness of prevastatin for the treatment of cardiovascular disease in hypercholester-olaemic men in Belgium, South Africa and Sweden	Cardiovascular disease (Prevastatin, usual care)	B, CA, SW SA	Cost per LYS	A, E, H, L, N	The effectiveness data (non-fatal MI, death) were taken from the WOSCOPS clinical trial. Resources were valued using country-specific costs and standardised to US dollars. An assumption was that the risk difference reduction derived from Scottish data would be applicable to other countries	Trial/model	Sweden = $8150/LYS (*R* ratio = 0.22); Belgium = $14,773/LYS (*R* ratio = 0.16); South Africa = $10,999/LYS (*R* ratio = 0.13)	The precise estimate does depend upon the specifics of the country, but the variations do not have much of an impact on the treatment decision that the study supports	I

continued

Study	Objective	Clinical area/ (technology)	Location	Outcome measure	Variation factor in *Table 39*	Methods	Vehicle	Conclusions	Other information	Dimension
Simpson[104]	To estimate the cost-effectiveness of adding zalcitabine to current antiviral treatment for AIDS patients with CD4+ counts of <300/μl	AIDS (Zalcitabine, zidovudine)	CH, DK, F, IT, UK	Cost per LYG	A, H	The effectiveness data (CD4+ cell response) were taken from an early clinical trial. Resource utilisation data were expert opinion and epidemiological data adjusted to reflect the AIDS profile in each country. Resources were valued using country-specific costs and standardised to ECUs (Euros, €)	Model	C/E ratios for zalcitabine were similar in the 5 countries studied: France = €13377/LYS, Germany = €17916/LYS, Italy = €12188/LYS, Switzerland = €15129/LYS, UK = €20708/LYS	The data regarding the standard AIDS treatments in each country were based on the expert opinion of a small group of physicians and it is possible that other physicians could specify other resource use patterns	I
Pinto[105]	To assess the cost-effectiveness of emedastine versus levocabastine in the treatment of acute allergic conjunctivitis	Acute allergic conjunctivitis Emedastine Levocabastine	B, D, F, NL, NO	Cost per additional SFD	A, E, N	The effectiveness data (failure and time without symptoms) were estimated from the clinical trial and combined with resource data (number of visits, drugs and laboratory examinations) established in each country from expert opinion	Model	The cost of failure was lower with emedastine than levocabastine in all European countries. Emedastine was economically dominant in Belgium, Germany, Portugal and Sweden. Germany = –€0.02; Portugal = –€5.31; Sweden = –€2.57. The incremental cost was low (<€1 per additional SFD) in France, The Netherlands and Norway. France = €0.83/SFD); The Netherlands = €0.09/SFD and Norway = €0.03/SFD	France has the highest ICER because the price of levocabastine is about 50% less than in other EU countries	I

continued

Study	Objective	Clinical area/ (technology)	Location	Outcome measure	Variation factor in *Table 39*	Methods	Vehicle	Conclusions	Other information	Dimension
Hakkaart-van Roijen[83]	To assess the cost-effectiveness of tapered versus abrupt discontinuation of a microemulsion formulation of cyclosporin in patients with plaque psoriasis	Plaque psoriasis Tapered discontinuation of cyclosporin Abrupt discontinuation of cyclosporin	CA, E, TU, UK	Cost per STFD	A, E, N	The effectiveness data (the number of STFDs) and resource utilisation data (medication, dermatologist visits, laboratory tests, GP visits, working days lost) were taken from a non-blind international RCT. Resources were valued using country-specific costs and converted to US dollars	Trial	Tapered discontinuation was dominant in Spain, Turkey and UK. Canada $1.4/STFD, Spain = $16.2/STFD, Turkey $104.6/STFD, UK = $–0.2/STFD	Relative price differences may have an impact on cost-effectiveness. The price of cyclosporin was lower in Spain. Therefore, it was considered difficult to pool the cost data. The importance of reporting results of different cost components is emphasised	I
Ghatnekar[102]	To assess the cost-effectiveness of becaplermin + good wound care compared with good wound care alone in the treatment of diabetic foot ulcers	Diabetic foot ulcers Becaplermin GWK GWK alone	CH, F, SW, UK	Cost per additional ulcer-free month gained	A, E, G, N, P	The effectiveness data (20-week healing rates) were taken from a meta-analysis of clinical trials, and transition probabilities were taken from a prospective study for the Markov model. Resource utilisation data/costs (inpatient care, topical treatment, outpatient care, medications, orthopaedic appliances) were based on high- and low-intensity resource patterns (Swedish and Swiss) and validated by expert opinion. Costs were converted to US dollars	Model	Becaplermin is cost-saving in Sweden, Switzerland and the UK. Sweden = –$384; Switzerland = –$279; UK = –$468. Becaplermin has an incremental cost per ulcer-free month in France. France = $142/ulcer-free month. Incremental costs of becaplermin treatment were lower with Swiss resource usage	Great differences in treatment practices and costs exist. Topic treatment costs are highest in Sweden and lowest in Switzerland. The low frequency of professionally administered dressing changes in Switzerland is related to the high labour costs and encouragement of patients to be self-reliant. Substitution effects are evidence between inpatient and outpatient settings	I

continued

Study	Objective	Clinical area/ (technology)	Location	Outcome measure	Variation factor in *Table 39*	Methods	Vehicle	Conclusions	Other information	Dimension
Jansen[93]	To assess the cost-effectiveness of continuous Lamisil (terbinafine) versus intermittent Itraconazole in the treatment of onychomycosis	Onychomycosis Terbinafine Itraconazole	D, I, IT, NL, SF, UK	Costs per complete cure	A, E, N	The effectiveness data (complete cure rates) were taken from the Lamisil versus Itraconazole in Onychomycosis (L.I.ON) study and resource utilisation data (routine medical management, adverse event management and management of relapse or failure) were established by expert opinion. Resources were valued using country-specific costs and reported in both local currency and US dollars	Model	Continuous terbinafine was dominant in Germany, Iceland, Italy, The Netherlands and the UK. Germany = –€173, Iceland = –€111, Italy = –€59, The Netherlands = –€84, UK = –€37. Continuous terbinafine has an incremental cost per additional complete cure in Finland. Finland = €524/additional complete cure	The study assumes that the only difference in medical management is the choice of antimycotic and the duration of treatment. However other studies suggest there may be inter-country differences associated with medical management (i.e. numbers of visits, frequency of laboratory tests). It is mentioned that effectiveness may differ from the efficacy results as observed in the L.I.ON study owing to patient selection, physician selection and compliance	I

continued

Study	Objective	Clinical area/ (technology)	Location	Outcome measure	Variation factor in *Table 39*	Methods	Vehicle	Conclusions	Other information	Dimension
Casciano[103]	To assess the cost-effectiveness of venlafaxine XR versus SSRIs and TCAs for the treatment of depression	Depression Venlafaxine XR SSRIs TCAs	CH, D, E, IT, NL, PO, SW, UK, US, V	Cost per STP Cost per SFD	A, E, N, Z	The effectiveness data (inpatient efficacy rates, outpatient efficacy rates, dropout rates due to lack of efficacy/adverse events) were taken from a meta-analysis of clinical trials, and resource utilisation data were established by expert opinion. Resources were valued using country-specific costs and converted to US dollars	Model	Venlafaxine XR was the most cost-effective in all countries except Poland. Germany = $16,257/STP; Italy = $11,590/STP; The Netherlands = $17,367/STP; Spain = $10,027/STP; Sweden = $10,951/STP; Switzerland = $15,398/STP; UK = $10,672/STP; US = $21,474/STP; Venezula = $33,920/STP; Poland = $1921/STP. Among the 9 countries in which Venlafaxine XR ranked first in expected cost per success, the difference in expected cost per success for Venlafaxine XR and the second ranking comparator ranged from $166 (Italy) to $2161 (US) for outpatients and from $494 (UK) to $3750 (US) for inpatients. Venlafaxine XR was most cost-effective in 9 countries as measured by cost per SFD for outpatients and inpatients. In these countries, the difference for venlafaxine XR and the second ranking comparator (SSRIs) ranged from $1.98 (Italy) to $18.45 (USA) for outpatients and from $6.15 (UK) to $32.83 (US) for outpatients	The study attempted to reflect the implication of drug selection for the average patient. However, significant individual patient variations and differences in practice patterns may compromise the generalisability. Additionally, the deterministic model design and the secondary data might introduce potential bias into the data analysis	I

continued

Study	Objective	Clinical area/ (technology)	Location	Outcome measure	Variation factor in *Table 39*	Methods	Vehicle	Conclusions	Other information	Dimension
Shear[94]	To assess the cost-effectiveness of terbinafine versus ciclopirox, clotrimazole, ketoconazole and micronazole for the treatment of tinea infections	Tinea infections Terbinafine versus ciclopirox, clotrimazole, ketoconazole and micronazole	AU, CH, D	Cost per DFD	A, E, N	The effectiveness data (efficacy rates, relapse rates) were based on a meta-analysis of clinical trials, and resource utilisation data (medications, physician visits, laboratory tests) were established from expert opinion. Resources were valued using country-specific unit costs and presented in local currency and US dollars	Model	Terbinafine compares favourably in terms of cost-effectiveness to other therapies. Austria = $6.39/DFD; Germany = $0.84/DFD; Switzerland = $1.09/DFD. The C/E ratios for other drugs ranged from $1.46/DFD (micronazole in Germany) to $9.13/DFD (clotrimazole in Austria)	Terbinafine is expected to be the most cost-effective choice in Austria, Germany and Switzerland, but there was no discussion of reasons for inter-country differences in C/E ratios	I
Ament[106]	To assess the cost-effectiveness of pneumococcal vaccination in the prevention of invasive pneumococcal disease	Pneumococcal disease (Pneumococcal vaccination, no vaccination)	B, E, F, SC, SW	Cost per QALY	A, E, L, N	The effectiveness data (incidence rates, mortality) were taken from a case–control study. Resource utilisation data (hospitalisations, vaccines) were established from expert opinion. Resources were valued using country-specific unit costs and standardised to ECUs (Euros, €)	Model	The C/E of pneumococcal vaccination varied considerably across the 5 countries: Belgium = €25,907/QALY, France = €19,182/QALY, Scotland = €14,892/QALY, Spain = €10,511/QALY, Sweden = €32,675/QALY	Results were sensitive to mortality rates and the incidence of invasive disease. Country differences were not related to real variations in the magnitude of disease occurrence but rather differences in surveillance systems and case ascertainment	I

continued

Study	Objective	Clinical area/ (technology)	Location	Outcome measure	Variation factor in *Table 39*	Methods	Vehicle	Conclusions	Other information	Dimension
Berger[97]	To assess the cost-effectiveness of pacilitaxel–cisplatin combination therapy versus standard cyclophosphamide–cisplatin regimen as first-line therapy for the treatment of advanced ovarian cancer	Advanced ovarian cancer Pacilitaxel–cisplatin combination Cyclophosphamide–cisplatin regimen	D, E, F, IT, NL, UK	Cost per LYS	A, E, N	The effectiveness data (response to therapy, progression-free survival, survival) were taken from a retrospective cohort study, and resource utilisation data (medication, hospitalisation, consultations, laboratory tests, investigations) were established by expert opinion. Resources were valued using US unit costs	Model	Pacilitaxel–cisplatin combination compared favourably in cost-effectiveness in all countries. Germany = $9362/LYS; Spain = $6395/LYS; France = $6642/LYS; Italy = $11,420/LYS; The Netherlands = $7796/LYS; UK = $6403/LYS	The study demonstrates some differences between European countries concerning the proportion between chemotherapy drug costs and hospitalisation costs. Hospitalisation costs are very high (75.7%) in the cyclophosphamide–cisplatin group whereas in the UK, hospitalisation costs are almost equivalent to treatment and laboratory/consultation costs (32.6% vs 21.1% vs 37.1%)	I
Lorenzoni[99]	To assess the cost-effectiveness of recombinant tissue plasminogen activator versus streptokinase for the treatment of acute MI	Acute MI (Recombinant tissue plasminogen activator, streptokinase)	D, IT, UK, US	Cost per LYS	A, B, E, H, N, P	The effectiveness data (mortality) were taken from the Global Utilisation of Streptokinase and Tissue Plasminogen Activator for Occluded Coronary Artery (GUSTO) trial. Resource utilisation data (drugs) were established by expert opinion. Resources were valued using country-specific costs and standardised to both US dollars and ECUs (Euros)	Model	The C/E ratio varied widely because of the wide difference in drug costs. Among a population of 1000 patients: D = $147,402/LYS, IT = $163,517/LYS, UK = $112,344/LYS, US = $221,053/LYS	The C/E in 3 European countries and the USA used the global results of the GUSTO trial. Even within the study, clinical efficacy differed between European and American patients. Further in the GUSTO population, a significantly high number of recombinant tissue plasminogen activator patients underwent bypass surgery or coronary angioplasty within 30 days, thereby removing high-risk patients from the analysis	I

continued

Study	Objective	Clinical area/ (technology)	Location	Outcome measure	Variation factor in *Table 39*	Methods	Vehicle	Conclusions	Other information	Dimension
Grover[101]	To assess the cost-effectiveness of simvastatin for diabetes patients with CVD and those without	Diabetes with CVD and without (Simvastatin)	CA, D, E, F, IT, UK	Cost per LYS	A, E, H, L, N	The effectiveness data (mortality, CVD events) were derived using the Cardiovascular Disease Life-Expectancy Model. Resource utilisation data (hospitalisations, physicians, outpatient, emergency services, drugs) were established using expert opinion. Resources were valued using country-specific costs and standardised to US dollars	Model	Primary prevention among diabetic patients was as cost-effective as secondary prevention among CVD patients in all countries. For 40-year-old men with diabetes, US = $6017/LYS, Canada = $3869/LYS, France = $3666/LYS, Germany = $6129/LYS, Italy = $4789/LYS, Spain = $1575/LYS, UK = $4249/LYS. For 40-year-old men with CVD, USA = $12,111/LYS, Canada = $7156/LYS, France = $7233/LYS, Germany = $10,790/LYS, Italy = $8850/LYS, Spain = $3815/LYS, UK = $7467/LYS	Healthcare costs associated with diabetes appear to be strongly associated with CVD	I

continued

Study	Objective	Clinical area/ (technology)	Location	Outcome measure	Variation factor in *Table 39*	Methods	Vehicle	Conclusions	Other information	Dimension
Neumann[107]	To discuss how new data can affect the results of prior assessments and how cost-effectiveness estimates evolve over time in the light of new experience	Motor vehicle airbags Statins ICDs	Various	Not specified	B, L, Q, W, X, Y, Z	Cost-effectiveness over time was explored by reviewing economic evaluations in a series of case studies	Model	The authors did not report any quantitative results over time, but trends. As an example, in the case of airbags, earlier economic studies had overestimated airbag effectiveness, overestimated baseline fatality rates and injury rates and underestimated seat belt rates	Reimbursement authorities should recognise the uncertainty of results. Models should be transparent, subject to peer review, verified to the degree possible and corroborated by other models. Uncertainty in model inputs/outputs should be explicit to permit specific inputs to be validated even if the entire model cannot be. Decisions/regulations about what empirical and experimental evidence to collect should be collected by value of information analysis	2

1, Location; 2, time.
A, Austria; AU, Australia; B, Belgium; CA, Canada; CH, Switzerland; D, Germany; DK, Denmark; E, Spain; F, France; GR, Greece; I, Iceland; IR, Ireland; IT, Italy; MA, Malawi; MO, Mozambique; NL, The Netherlands; NO, Norway; NZ, New Zealand; P, Portugal; PO, Poland; SA, South Africa; SC, Scotland; SF, Finland; SW, Sweden; TA, Tanzania; TU, Turkey; UK, United Kingdom; US, United States; V, Venezuela; EU, European Union.
A, Absolute/relative costs (prices); B, artificial study conditions; C, capacity utilisation; D, case-mix; E, clinical practice variation; F, compliance; G, culture/attitudes; H, demography; I, disease interaction; J, opportunity cost; K, economies of scale; L, epidemiology; M, exchange rates; N, geographical setting; O, health state valuations; P, healthcare resources; Q, healthcare system; R, historical differences; S, incentives; T, industry-related bias; U, joint production; V, perspective; W, skills/experience; X, technological innovation; Y, timing of assessment; Z, treatment of comparators.
C/E, cost-effectiveness; DFD, disease-free day; DPA, disease progression avoided; GI, gastrointestinal; GORD, gastro-oesophageal reflux disease; ICST, inhaled corticosteroid; LYG, life-years gained; LYS, life-years saved; OA, osteoarthritis; RTI, respiratory tract infection; SFD, symptom-free day; SSRI, selective serotonin reuptake inhibitor; STFD, systemic therapy-free day; STP, successfully treated patient; TCA, tricyclic antidepressant; UGI, upper gastrointestinal.

(C) Papers suggesting methods to assess variability

Study	Objective	Systematic review methods	Vehicle	NHS-specific	Methods	Empirical example	Dimension
Drummond[31]	To use modelling to compare the cost-effectiveness of the treatment of acid-related disease across countries	No	Model	No	Using the same structural model and core assumptions, which are then applied to the specific contexts of four health care systems: Belgium, France, United Kingdom and the United States. All countries took core efficacy data from a US trial, but compliance with therapy, detection rates and consequent use of healthcare resources were allowed to vary by country. Cost comparisons between countries were made by adjusting to US dollars using purchasing power parities	Yes	I
Menzin[116]	To explore how data from pivotal clinical trials in one setting can serve as a basis for economic evaluations in others using the example of recombinant human deoxyribonuclease to improve pulmonary function in patients with cystic fibrosis	No	Model	No	Modelling based on the assumption that the relative risk reduction is generalisable to other settings, but resource utilisation is country-specific. Information is ascertained through expert opinion on 'practice pattern' parameters believed likely to vary across countries such as the likelihood of hospitalisation and the associated mean length of stay in hospital. A decision model was used as a means of extrapolating from the US trial data to the other locations	Yes	I
Postma[65]	To inform multinational HIV/AIDS public-health strategy planning in the European Union by integrating national studies on HIV/AIDS	No	Model	Yes	To estimate the economic impact of HIV/AIDS, an epidemic was simulated using two models (MIDAS and PC-based AIDS scenarios). Parameter values representing economic impact per patient-year were derived from a structured review of publications on economic aspects of HIV/AIDS. Cost data from existing studies can be used to produce preliminary and illustrative projections of AIDS impact on hospitals, but standardisation of these estimates must be improved if they are to be more valuable	No	I

continued

Study	Objective	Systematic review methods	Vehicle	NHS-specific	Methods	Empirical example	Dimension
Tolley[47]	To review several European HIV-AIDS cost studies and outline the differences in methods used and where better approaches to cost estimation could have been adopted	No	Not specific	No	To enhance the validity of comparing the costs of treatment and care packages across countries with different healthcare systems, a standardised cost framework is proposed. This allows the separation of actual (rather than methodological) variations in costs across locations. The standardised cost framework includes a checklist of standard cost categories that could be used in a cost study, a standardised approach to service use data collection and standard principles for estimating the unit costs of service use	No	I
Davis[149]	To argue that additional information is necessary to determine the extent of extrapolation or generalisation to other populations warranted in a specific clinical trial	No	Trial	No	To judge the extent to which the results of a clinical trial can be generalised requires additional information, which may be derived from basic science laboratory studies, animal studies, genetic studies (if applicable), observational, clinical and epidemiological studies and other randomised clinical trials with similar settings or treatments using a checklist framework. Two levels of generalisation exist: mechanistic and economic. The mechanistic question is 'Would the same pathological mechanism of a therapy apply in other segments of the population? and the economic question is 'Would the treatment be cost-effective in populations other than those studies in the trial?	No	I
De Jonghe[98]	To assess the cost-effectiveness of tuberculosis chemotherapy and to develop reasonable estimates of the marginal and average cost per case cured that could be of use to other countries at differing levels of income and infrastructure	No	Observa-tional	No	Generalising results to other developing countries is questionable because of variation in national income per capita. In order to construct more realistic estimates of the average incremental unit for treatment in countries with higher incomes, the component of each unit cost that is expenditure on local goods and services and the component that must be purchased from abroad in foreign currency should be distinguished. Cost can be divided into the external component denominated in dollars and the domestic component calculated in terms of percent of GDP per capita	Yes	I

continued

Study	Objective	Systematic review methods	Vehicle	NHS-specific	Methods	Empirical example	Dimension
Rovira[160]	To criticise the present variability of economic appraisal methodologies and support their standardisation as a means of ensuring the comparability of the results of the studies done by different researchers	No	Not specific	No	Establishing a comprehensive set of methodological norms, which operationally define the procedures for performing a given type of analysis, including: (1) definition of concepts and methodology; (2) criteria for the selection of alternatives to be compared; (3) establishment of the effects to be included in different types of analysis; (4) establishment of precise procedures for measuring, valuing and adjusting the effects considered; (5) specification of the formulae applied to aggregate the various effects in a single index; (6) presentation of the results in a form that ensures the transparency and replication of the analysis	No	I
Glick[111]	To consider issues in the design and planning of an economic assessment, including factors that influence the generalisability of the study's findings	No	Trial	No	Performing subanalyses of the clinical and economic data for a selected set of practice styles (e.g. separate analyses of costs and effects might be performed for centres with hospital lengths of stay that were longer than the average observed in the trial as well as for centres with shorter average stays). Two other approaches described include: (1) comparing trial-wide outcomes with trial-wide utilisation valued using centre-specific prices; (2) comparing trial-wide outcomes with centre-specific utilisation valued using centre-specific prices	No	I
Oster[128]	To compare the cost-effectiveness of two alternative approaches to the hypercholesterolaemia treatment in a RCT by increasing its generalisability to general practice	No	Trial	No	To enhance generalisability, the trial contained the following characteristics: (1) representative models of care; (2) compliance with plans of treatment was encouraged, but not enforced; (3) follow-up of 1 year irrespective of remaining on therapy; (4) minimal external monitoring; (5) costs of medications borne by patients who may insist on alternatives to their randomised treatment; (6) stratification of costs according possession of insurance coverage	No	I

continued

Study	Objective	Systematic review methods	Vehicle	NHS-specific	Methods	Empirical example	Dimension
Heyland[26]	To determine the extent to which economic evaluations published in the critical care literature provide information that can help to improve the efficiency of an intensive care unit	No	Not specific	No	Using criteria to assess the generalisability of economic evaluations based on: (1) clinical generalisability; (2) systems generalisability. Clinical generalisability relates to similarity of patients to those in user's setting. Systems generalisability relates to appropriateness of viewpoint, availability of intervention, costing methods, unit prices, mix of resources, volume of patients, exchange rates, outcome measurement, patient preferences and discount rates to user's setting	No (but review of published studies exists)	I
Jefferson[140]	To discuss the problems of reviewing available economic information and to explore the possibility of summarising the results of a review in terms of quantities of resources, rather than prices	No	Model	No	Performing secondary analysis based on studies identified through a systematic literature search and screened against inclusion criteria. Cost estimates are stratified by age and for each population group, the 'benefit-cost ratios' are listed and the distribution of the estimates identified into the 25th, median, 75th percentile values ('low', 'medium' and 'high'). All costs quoted in the studies are extracted and standardised to 1987 US dollars based on official exchange rates and PPPs. Quantities of resources are extracted to which a standard set of unit costs were applied. These are used in a series of evaluations performed on a simulated population cohort of 1 million individuals, holding the efficacy at 80% to generate a series of 'benefit-to-cost ratios'	Yes	I
Jefferson[146]	To determine whether scientific reviewing methods can be adapted to economic studies, to sum up evidence about the cost and cost-effectiveness of influenza vaccine and to carry out an experimental secondary economic evaluation	No	Model	No	Constructing secondary analysis using a resource costing approach, which considers the possibility of deriving data on resource inputs from existing studies, and estimating costs and cost-effectiveness from unit cost specific to a particular setting. Ranges of estimates were considered and pooled into an economic model based on a simulated general population cohort of 1 million individuals. Pooling consisted of construction of a distribution of resource estimates to which marginal unit costs of hospital, other care and lost working time were derived from a cost of illness study. Robustness of 'benefit-to-cost ratios' to changes in costs and influenza attack rates were assessed in sensitivity analyses, fixing the efficacy of the vaccine throughout	Yes	I

continued

Study	Objective	Systematic review methods	Vehicle	NHS-specific	Methods	Empirical example	Dimension
Rittenhouse[43]	To question whether conventional efficacy measurements in the form of RCTs are appropriate for answering questions of interest to economists	No	Model	No	Using modelling to modify the results from RCTs to predict performance better in the real world of clinical practice based on conjecture or prior experience	No	I
Schulman[77]	To explore methods related to analysis of the economic evaluation alongside the Flolan International Randomised Survival Trial (FIRST) study	No	Trial	No	To sample the country-specific health production functions in all study countries requires assessment of resources used to care for patients, the outcomes resulting from the use of these resources and the incremental effect of the treatment under study on costs and effects of therapy. In a trial-specific analysis, a single study resource cost for each item is based on the weighted average of the individual country costs. Country-specific analyses would apply single-country unit costs to aggregate levels of resources consumed by trial patients. This type of analysis would potentially be more usefully for local country decision-makers	No	I
Drummond[1]	To assist the user in determining whether the results of an economic study are valid	No	Not specific	No	Using a checklist approach to determine whether an economic analysis yields valid and important results which can then be applied to the user's clinical setting based on (1) validity of the results; (2) magnitude of the results and (3) implementation of the results	No	I
Jonsson[33]	To address the problems that arise in economic evaluations linked to international trials using as an example the economic substudy of GUSTO IIb	No	Trial	No	To assess the representativeness of baseline data on resource utilisation in different countries, a pooled proportional difference in resource utilisation across countries can be calculated and then applied to country-specific baseline costs. The validity of modelling life expectancy to countries outside the USA can be checked using methods such as comparing risk functions for a secondary infarction in the placebo group in the Scandinavian Simvastatin Survival Study and the associated survival functions after a secondary infarction	No	I

continued

Study	Objective	Systematic review methods	Vehicle	NHS-specific	Methods	Empirical example	Dimension
Rice[21]	To explore the use of MLM in health services research	No	Trial	No	Suggesting the application of MLM to detect centre-related differences in RCTs as outcome may be less dependent on the treatment received, but centre characteristics	No	I
Manning[143]	To identify threats to both the internal and external validity of estimated treatment effects and suggest possible econometric and experimental solutions	No	Trial Observational	No	To suggest the use of econometric and experimental solutions for addressing bias in observational and RCT studies to improve estimates of treatment effects and incremental costs. The primary objective of evaluation is to provide information to guide future decisions concerning patients other than study participants, but because the latter may differ from future classes of patients in terms of unobservable characteristics, selectivity is as important a threat to external validity as it is to internal validity. Alternative designs should be viewed as differences of degree (the validity of instruments for selection and participation) rather than kind (whether there was random allocation within the study)	No	I
McClellan[144]	To explore the use of instrumental variables for estimating the marginal cost-effectiveness of a medical technology	No	Observational	No	Using panel instrumental variable estimation methods to obtain estimates of cost and outcomes in observational datasets that eliminate the potentially large bias resulting from unobserved heterogeneity	Yes	I
Rittenhouse[139]	To show how one may supplement a trial with additional data to connect the artificial trial to the real world of clinical practice	No	Trial	No	To propose supplementing trial data with additional data to connect the artificial trial to the real world of clinical practice through modelling. If data are not directly available, a Bayesian approach to revising trial estimates could be adopted. For example, one trial measure which might be adjusted is the proportion of patients with disease who would have presented with clinical signs if they had not been part of a trial that allowed early detection and treatment based on subclinical testing mandated by the trial protocol. In routine practice, those presenting with clinical signs would use additional resources for treatment/and or confirmatory diagnostics; those with subclinical disease would either never use resources if they never developed clinical manifestations, or use resources of a different type or intensity if perhaps they have different outcomes by virtue of their disease being discovered at a later time. Basing resource use and ultimate effectiveness on the revised measure of outcome rather than one in the trial could lead to more accurate estimates of the cost-effectiveness of interventions in routine practice	No	I

continued

Study	Objective	Systematic review methods	Vehicle	NHS-specific	Methods	Empirical example	Dimension
Coyle[127]	To discuss one of the threats to generalisability: the existence of protocol-driven costs	No	Trial	No	Discusses a number of approaches including (1) modelling based on efficacy results from RCT data and data outside the trial; (2) pragmatic trials; (3) changes to data collection within RCTs by adapting study designs to allow analysis of costs which have only occurred for clinical reasons; (4) modifying protocols in select centres to examine effects on resource consumption and to design data collection such that contact is limited to that which would occur in real practice between the study physician and the patient	No	I
Haycox[115]	To test the international transferability of a disease model for UGI disease to different countries	No	Model	No	Internationalising a model consists of three steps: (1) Identification of variations in national cost-containment policies with their subsequent impact on clinical practice and resource utilisation; (2) measurement of the impact of such variation on health and economic outcomes; (3) valuation of the extent to which such variations altered the global costs and effectiveness for UGI disease	Yes	I
Johnston[125]	To explore the generalisability of costs from a trial investigating the frequency of breast cancer screening in the UK	No	Trial	Yes	Exploring whether trial centres are representative of the other centres in terms of organisational characteristics and adjusting costs for any biases inherent in the organisation of trial centres through the use of sensitivity analysis	Yes	I
Kinosian[141]	To review the key issues to be considered when extending trial data to real-world situations	No	Trial	No	Developing risk profiles to identify a population in a clinical trial and then to extend data for this population to subjects outside the trial with the same set of risk factors. Two key issues must be considered: (1) the importance of cholesterol history over time and (2) the selection of the measure to be used to stratify the population and impute effects	No	I

continued

Study	Objective	Systematic review methods	Vehicle	NHS-specific	Methods	Empirical example	Dimension
Mugford[126]	To estimate costs and cost-effectiveness of a new model with a traditional antenatal care package in Argentina, Cuba, Saudi Arabia, Thailand and explore whether trial economic results are generalisable to other lower-resourced settings	No	Trial	No	Investigating the feasibility of generalising cost and cost-effectiveness evidence from trial centres to non-trial centres in South Africa, Gambia, Zimbabwe, Indonesia and Bangladesh. From non-trial data including local characteristics and features of the health service including prices, morbidity and utilisation patterns, investigations can be made of whether there are enough resources to adopt either programme of care and whether the skills are available to deliver as effectively as in the trial	No	I
Rutten-van Molken[133]	To determine the relative economic consequences of treating asthmatic subjects with twice daily dry powder formoterol 12 μg as compared with salmeterol 50 μg after correction for country differences	No	Trial	No	Using regression analyses to correct for possible differences between countries and patients when testing the impact of the trial medication on total costs. Differences in definitions, costing methodology and uncertainty associated with data sources explain part of the variation in unit prices so it is difficult to pool the cost data from the 6 participating countries. Therefore, the importance of separating resource use from costs rather than simply reporting and analysing total costs is emphasised	Yes	I
Schulman[120]	To describe a framework for the collection and analysis of international resource cost data that can contribute to a consistent and accurate inter-country cost-estimation	No	Trial	No	Developing a standardised costing methodology in seven countries by (1) converting total average costs to approximations of marginal unit costs and (2) converting costs from different countries into a common currency using purchasing power parities for the study countries. Where there were still several resource items for which unit costs were unavailable for certain countries, analysts developed an index table, based on a market-basket approach using a method based on physician-work and practice-expense resource-based relative value units	Yes	I

continued

Study	Objective	Systematic review methods	Vehicle	NHS-specific	Methods	Empirical example	Dimension
Willke[25]	To explore a method for evaluating the generalisability of the overall cost-effectiveness results from a multinational clinical trial to individual countries participating in the trial for individual countries	No	Trial	No	Using regression models that take into account the treatment that a patient receives, a series of exogenous variables (e.g. disease severity), the health outcome of the episode and country-by-treatment and country-by-outcome interaction terms. The model specification allows treatment to affect cost independent of outcome, and allows the outcome to affect the cost of care independent of treatment to permit customisation of cost-effectiveness ratios	Yes	I
Baltussen[114]	To propose a 3-step approach, comprising successive assessment of internal validity, external validity and net impact at the system level to enhance the informative value of economic analysis	No	Not specific	No	Using a three-step approach: step 1 involves deriving internally valid results through collecting effectiveness parameters and patient-specific data on resource utilisation in an RCT; step 2 involves adapting the results to the real world through taking into account context-specific factors such as the various procedures included in the economic evaluation, and specific physician, hospital and healthcare system characteristics; step 3 implies the assessment of costs and outcome changes associated with introducing the new intervention	No	I
Caro[130]	To generalise the results of WOSCOPS to the perspective of any national health service or other organisation responsible for societal costs	No	Trial/model	No	Deriving a generalised formula expressing the ratio in terms of elements that might be country-specific and those that ought to be more general. It is assumed that the relative treatment effect is generally applicable. However, the baseline risk can be affected by differences in the distribution of risk factors in different populations. With this model, one can calculate the cost-effectiveness for any given baseline risk. The major factor that clearly characterises the country is the cost structure, but full accounting of these details would require separate models	Yes	I

continued

Study	Objective	Systematic review methods	Vehicle	NHS-specific	Methods	Empirical example	Dimension
Dixon[112]	To illustrate how economic evaluation alongside TARGET has been designed to avoid threats to external validity that an awareness of clinical, operational and economic aspects of a service can reveal	No	Trial	No	To provide more valid data on resource use and costs requires the use of a number of approaches. Simpler problems (e.g. unit cost bias) can be explored using one-way sensitivity analysis using routine cost data. More complex problems can be dealt with by use of observational designs: (1) a prospective observational study; (2) the GPRD database; (3) a retrospective study; (4) a prospective study; (5) analysis of hospital tariffs across the UK as a useful measure of true cost variability and to inform sensitivity analysis	No	I
Goeree[53]	To develop a conceptual framework for selecting hospitals for unit cost estimates in national and international multicentre trials and to test the impact of alternative hospital selection on cost results	No	Trial	No	To assess the generalisability of unit cost estimates, three considerations must be taken into account: (1) the number of hospitals chosen for unit cost estimates; (2) the sampling method for selection of hospitals; (3) desired level of subgroup analysis by geographical area	Yes	I
Grieve[84]	To test the validity of the assumption of homogeneity of costs between countries in multinational economic evaluations	No	Trial	No	To illustrate that costs measured in one country may not be generalisable through the collection of country-specific costs. For each patient, durations of hospital stay, use of investigations, average medical time and local prices were collected and directly compared. Pooled data used in multinational economic evaluations may not be applicable to the local setting and instead evaluations should report disaggregated resource use data and county-specific costs which would be more useful to the decision-maker	Yes	I

continued

Study	Objective	Systematic review methods	Vehicle	NHS-specific	Methods	Empirical example	Dimension
Pang[137]	To suggest the application of cluster analysis and MLM to multinational economic evaluation data	No	Trial	No	To investigate patterns in resource utilisation between different countries or centres and examine the use of MLM to detect explicitly a hierarchical structure to multinational economic evaluation data if one is present and to explore the relationship between centre and outcome. The presence of clustering can have important implications for costing methodology and pooling of economic evaluation data	No	I
Rutten-van Molken[134]	To compare methods for aggregating resource utilisation and cost data from different countries	No	Trial	No	To pool data, two approaches were used: (1) applying country-specific unit prices in national currencies to country-specific volumes of resource use and then converting to US dollars; a regression equation is estimated using dummy variables to control for differences between countries; (2) pooling volumes and adjusting these volumes to reflect national medical practice better and then applying country-specific unit prices in national currencies to the pooled volumes	Yes	I
Shih[142]	To overcome some of the bias inherent in RCT data while avoiding some of the common pitfalls associated with the use of observational data	No	Model	No	Using a decomposition technique between different population subgroups based on the principle of removing population effects from observed outcome effects and identification of the separate components of the outcome effect (treatment and population) is key. The treatment effect is the difference in relative risk applied to the treatment group, whereas the population effect is an adjustment for differences in 2 study populations. Multivariate regression models are used to regress costs on sets of risk factors such as age, gender, race, disease group and logistic/linear regression used to estimate branch probabilities and payoffs within a decision-modelling framework	Yes	I
Brennan[158]	To itemise the current and developing roles of modelling in economic evaluation and to discuss its value in each role	No	Model	No	Using modelling to bridge between efficacy and effectiveness and for tailoring of a robust set of clinical results to the different modes of organisation in particular locations	No	I

continued

Study	Objective	Systematic review methods	Vehicle	NHS-specific	Methods	Empirical example	Dimension
Evans[157]	To examine intrastudy effects on validity	No	Trial	No	Creating data collection strategies according to the level of internal validity found in an economic evaluation. Instead of viewing internal and external validity as polar opposites, validity should be considered in terms of a continuum within a particular study. The use of proxies to collect resource utilisation estimates, the reliance on patient self-reported data and the method of collecting this type of data all impact the validity of study results	No	I
Murray[29]	To outline methods of CEA, the limitations of current methods, directions for revising these methods and some of the remaining technical challenges facing this revision	No	Not specific; model by implication	No	Proposing a modification of the standard intervention current mix (ICM)–CEA lifting the constraint on the current mix of interventions to evaluate the cost-effectiveness of all options including currently funded interventions. These modifications include: (1) evaluating the costs and benefits of a set of related interventions with respect to the counterfactual of the null set of the related interventions; (2) presenting the results of CEA in a single league table. The decision rules are similar to those that have been derived for intervention mix constraint (IMC)–CEA, but the analysis starts from the origin by relating cost-effectiveness to 'do nothing' options	No	I
Raikou[119]	To assess whether the use of average or centre-specific unit costs in multicentre trials have an impact on average treatment costs	No	Trial	No	Comparing two methods for collecting costing information within a simulated clinical trial setting (1) estimating average treatment costs by applying unit costs averaged across treatment centres to centre-specific volumes of resource use, (2) using centre-specific information for both the unit costs and the resource volumes, and then averaged across centres. Using a prespecified production relation between the different volumes of resource use, and simulating changes in unit costs, the two methods result in statistically different estimates of average treatment costs. Therefore a more cautious approach should be adopted in the collection, calculation and interpretation of treatment costs in multicentre studies	No	I
Sorensen[136]	To identify sources of cross-national variation in the cost of an episode of care, approaches to adjust variation and practical problems in trans-national adaptation of economic clinical trials	No	Trial	No	To address sources of trans-national variation include (1) substituting unit prices of the target country, (2) adjusting observed resource quantity to typical treatment pattern of the target country, (3) replacing selected treatment encounters with functionally analogous service existing in target country and (4) subsetting trial cohort to obtain outcomes in a patient population representative of the target population	No	I

continued

Study	Objective	Systematic review methods	Vehicle	NHS-specific	Methods	Empirical example	Dimension
Tunis[129]	To review some of the methodological challenges in the design of medical effectiveness trials for olanzapine	No	Trial	No	Enhancing the external validity of costs and effectiveness through naturalistic designs by relaxing several elements: (1) physicians are allowed to adjust dosages; (2) heterogeneous patient population with a variety of comorbidities including substance abuse; (3) administering treatment in an open-label fashion and flexibility in the use of non-pharmacological supportive and rehabilitative therapy	No	I
Cook[131]	To explore the use of statistical methods for testing homogeneity of clinical and economic treatment effects among countries in the Scandinavian Simvastatin Survival Study (4S)	No	Trial	No	Evaluating whether pooling of economic data is appropriate through the conduct of homogeneity tests to detect country-by-treatment interactions. A qualitative or crossover interaction occurs when the treatment effect is positive for patients in some countries and negative in others. A non-crossover interaction occurs when the magnitude, but not the direction, of treatment varies. Advantages of using net benefit measures relative to ICERs are also described	Yes	I
Coyle[138]	To assess the degree of variation in costs between patients and treatment centres and the determinants of such variation	No	Trial	No	Using one-way ANOVA to identify the degree of variation in costs between centres within each subgroup, and using OLS regression to assess the degree to which variation could be explained by various factors relating to unit costs, patient characteristics and centre characteristics	Yes	I
Glick[121]	To investigate for collection of hospital unit cost data for use in multinational clinical trials, for how many types of hospitalisation and how many countries should estimates be obtained	No	Trial	No	To investigate whether reliable methods for imputing hospitalisation costs exist to allow one to economise on data collection, subsets of hospitalisation types were randomly sampled and used to develop imputation regressions	Yes	I

continued

Study	Objective	Systematic review methods	Vehicle	NHS-specific	Methods	Empirical example	Dimension
Hutton[132]	To gain an understanding of the differences between healthcare settings that reduce the accuracy of cost generalisations	No	Trial	No	To demonstrate that cost generalisations within and between countries are unreliable through the collection and analysis of cost data in a multinational trial of antenatal care practice. Prices, resource use and health service use were compared between women with different case mix, health facilities, trial arms and study countries. Generalisation is only possible if levels of resource productivity, occupancy levels, staffing patterns, prices and exchange rates, input mix and health facility size are similar between settings	No	I
Spath[147]	To propose a three-step approach to evaluate the potential of economic studies for transfer	Yes	Not specific	No	Using a checklist approach to identify relevant economic evaluations based on four inclusion criteria: perspective, number of competing options, appropriate therapies and relevance of therapies to the healthcare system concerned. This is followed by an analysis of the eligibility for transfer based on five indicators based on three dimensions: settings in which the studies might be used, the transferability of health outcome data and the transferability of resource utilisation data	Yes	I
Oostenbrink[159]	To use Markov modelling to convert trial-based cost-effectiveness information from one country to another	No	Model	No	Using Markov modelling to facilitate the transfer of trial-based data to other countries by modelling the difference between treatment groups observed in trials by monthly transition probabilities between disease states and the probabilities of experiencing an event in a given disease state. As resource use is linked to disease states and events, a Markov model can be used to incorporate country-specific resource use and unit costs	Yes	I
Welte[148]	To propose a checklist of transferability factors to assess the transfer of economic evaluation results between countries	Yes	Not specific	No	Using a checklist that groups transferability factors into methodological context, healthcare system and population characteristics. Different levels of effort are required for the analysis ranging from very low (e.g. discount rate) to very high (e.g. staff characteristics). The transfer of study results may require adjustment for time (using price indices), different currencies (using PPPs), effects, resource utilisation and monetary revaluation. The checklist enables a quick determination of possible transfer problems, the identification of the most needed adjustments and 'knock-out' criteria (e.g. great difference in disease incidence)	No	I

continued

Study	Objective	Systematic review methods	Vehicle	NHS-specific	Methods	Empirical example	Dimension
Koopmanschap[135]	To investigate how costs should be handled in multinational economic evaluation	No	Trial	No	Using regression analyses to correct for differences in medical practice and calculating the costs of medical consumption as if all patients were treated in one country. Corrected amounts of resource consumption can be multiplied with country-specific costs	Yes	1
Ludbrook[122]	To investigate the conversion of prices to a common base in order to make results comparable in economic evaluations across countries	No	Trial	No	Using three ways to convert costs to a common currency base: (1) simple exchange rates, (2) healthcare PPP and (3) specific PPP for dialysis. Conversion factors can affect the interpretation of economic evaluation results particularly where input mix and relative prices vary significantly	No	1
Sullivan[124]	To present methodological options for conducting economic evaluations alongside multinational clinical trials	No	Trial	No	Using two analytical approaches: (1) pooling all patients with unit costs from one country and altering costs in a sensitivity analysis, (2) multivariate statistical analysis controlling for regional or country differences in the supply of healthcare, healthcare utilisation and patient behaviour using GNP, adjusted by purchasing power parity	No	1

1, Location; 2, time.

(D) Empirical application of methods to increase generalisability in economic evaluation

Study	Objective	Clinical area	Methods	Vehicle	Conclusions	Other information	Dimension
Drummond[31]	To use modelling to compare the cost-effectiveness of the treatment of acid-related disease across countries	Osteoarthritis (Misoprostol)	Using the same structural model and core assumptions, which are then applied to the specific contexts of four healthcare systems: Belgium, France, the UK and the USA. All countries took core efficacy data from a US trial, but compliance with therapy, detection rates and consequent use of healthcare resources were allowed to vary by country. Cost comparisons between countries were made by adjusting to US dollars using PPPs	Trial	The UK made the greatest use of ambulatory care, probably because of relatively well-developed primary care services in that country. The acquisition cost of misoprostol was similar for the three European countries, but was higher in the USA. The four countries differed both in the percentage of patients who were hospitalised with ulcer and the percentage who underwent a surgical operation. Duration of hospital stay was shortest in the USA, but the cost of surgical care was by far the highest. The estimated net costs of 3 months of misoprostol prophylaxis showed surprisingly similar results, despite differences between the countries in clinical practice patterns and cost components	The main problems that this study encountered related to the availability and type of data found in each country, a common problem in modelling studies	I

continued

Study	Objective	Clinical area	Methods	Vehicle	Conclusions	Other information	Dimension
De Jonghe[98]	To assess the cost-effectiveness of tuberculosis chemotherapy and to develop reasonable estimates of the marginal and average cost per case cured that could be of use to other countries at differing levels of income and infrastructure	Tuberculosis (Chemotherapy)	Generalising results to other developing countries is questionable because of variation in national income per capita. In order to construct more realistic estimates of the average incremental unit for treatment in countries with higher incomes, the component of each unit cost that is expenditure on local goods and services and the component that must be purchased from abroad in foreign currency should be distinguished. Cost can be divided into the external component denominated in dollars and the domestic component calculated in terms of percent of GDP per capita. Detailed cost data were collected from the National Tuberculosis Programmes (Tanzania, Malawi and Mozambique) during the first 6 months of 1990. Costs included programme management, laboratory, drugs, hospitalisation and ambulatory treatment. Sources included accurate budget sheets and accounting books, National Reference Laboratory reagent and equipment inventories, salary scales and receipts. Programme benefits were estimated using data from a cohort study and a transmission model	Observational	Costs per patient treated for the different regimens in Malawi, Mozambique and Tanzania are remarkably similar except for the differences in costs of food and labour. Short-course chemotherapy is cheaper than standard chemotherapy per death averted and year of life saved for both hospital and ambulatory strategies, except for the marginal cost of ambulatory chemotherapy	Authors conclude that these are fairly robust conclusions. Desirability of hospitalising patients to increase compliance depends on locally specific determinants and marginal cost of hospitalisation, hence generalisations about the role of hospitalisation are not possible	I

continued

Study	Objective	Clinical area	Methods	Vehicle	Conclusions	Other information	Dimension
Jefferson[140]	To discuss the problems of reviewing available economic information and to explore the possibility of summarising the results of a review in terms of quantities of resources, rather than prices	Influenza (Vaccination)	Performing secondary analysis based on studies identified through a systematic literature search and screened against inclusion criteria. Cost estimates are stratified by age and for each population group, the BCRs are listed and the distribution of the estimates identified into the 25th, median, 75th percentile values ('low', 'medium' and 'high'). All costs quoted in the studies are extracted and standardised to 1987 US dollars based on official exchange rates and PPPs. Quantities of resources are extracted to which a standard set of unit costs were applied. These are used in a series of evaluations performed on a simulated population cohort of 1 million individuals, holding the efficacy at 80% to generate a series of BCRs	Model	Benefit-to-cost ratios change with different resource estimates and attack rates. There was an inverse relationship between incidence/attack rates and benefit-to-cost ratios	Methodological issues arise when attempting to perform 'secondary analysis' of existing economic evaluations including: (1) terminology (i.e. definition of economic evaluation; (2) aim of the review; (3) distinction between primary and secondary economic evaluation; (4) standardisation of costs; (5) specification of healthcare context; (6) quantities of inputs and variations in combinations of factor inputs; (7) standardisation of price, currency and technology levels. Developmental progress may require a more coherent theoretical framework linked to cost and production function theory	I

continued

Study	Objective	Clinical area	Methods	Vehicle	Conclusions	Other information	Dimension
McClellan[144]	To explore the use of instrumental variables for estimating the marginal cost-effectiveness of a medical technology	MI (Intensive care procedures)	Using panel instrumental variable estimation methods to obtain estimates of cost and outcomes in observational datasets that eliminate the potentially large bias resulting from unobserved heterogeneity. Data on the treatment of heart attacks in the US elderly population from 1987 to 1990 were used to estimate panel IV models of the incremental costs and outcome effects of intensive cardiac procedures [cardiac catheterisation, subsequent revascularisation (angioplasty or bypass surgery)]. Indicator variables for the technological capabilities at each hospital over time were based on the date and quantity of invasive procedures performed on elderly acute MI patients. Difference-in-differences estimation methods were used to identify treatment effects with an emphasis on testing the validity of identifying assumptions. To obtain IV estimates of the outcome and cost effects of the incremental treatments resulting from technology adoption, models were implemented that estimated weighted-average treatment effects over all demographic groups and all adopter-non-adopter hospital DD comparisons. A saturated analysis of variance model with binary interactions for patient demographics and a full set of hospital fixed effects and year fixed effects to remove average differences in treatments and outcomes associated with these observable factors were estimated. Interactions between hospital type and year as instrumental variables were used to estimate residual DD treatment and outcome relationships	Observational	The most conservative cost-effectiveness estimate for the diffusion of invasive acute MI technologies is at least $40,000 per patient surviving to 1 year. If some of the mortality reduction is attributed to the timing of adoption decisions after bad years, the cost-effectiveness estimate is ~$70,000. By contrast, least-squares estimates based on adjustments for observable patient differences would lead to a cost-effectiveness ratio under $30,000	This study involves the average cost-effectiveness of changes in medical treatment patterns rather than the population average cost-effectiveness for an observable population and therefore is likely to be more relevant for policy analysis	I

continued

Study	Objective	Clinical area	Methods	Vehicle	Conclusions	Other information	Dimension
Jefferson[146]	To determine whether scientific reviewing methods can be adapted to economic studies, to sum up evidence about the cost and cost-effectiveness of influenza vaccine and to carry out an experimental secondary economic evaluation	Influenza (Vaccination)	Constructing secondary analysis using a resource costing approach, which considers the possibility of deriving data on resource inputs from existing studies and estimating costs and cost-effectiveness from unit cost specific to a particular setting. Ranges of estimates were considered and pooled into an economic model based on a simulated general population cohort of 1 million individuals. Pooling consisted of construction of a distribution of resource estimates to which marginal unit costs of hospital, other care and lost working time were derived from a cost of illness study. Robustness of BCRs to changes in costs and influenza attack rates were assessed in sensitivity analyses, fixing the efficacy of the vaccine throughout	Model	The number of doctor–patient consultations and work-days lost do not appear to substantially affect the BCR, whereas the model is sensitive to the estimate used for length of hospital stay. The BCRs range from 0.3 to 1.2 for the low attack rate scenarios and from 1.1 to 3.6 for the high attack rate scenarios. There appears to be a direct relationship between estimated attack rates and BCR	The study does not prove conclusively that systematic review-based approach to economic studies can yield results that can be generalised to other countries. Other questions to be addressed for generalising the model are (1) how to specify the healthcare context; (2) how to standardise price levels and currency; (3) how to deal with differences in technology levels; and (4) how to adjust the model for differences in resource use and valuation and hence marginal costs of technologies	I
Johnston[125]	To explore generalisability of costs from a trial investigating the frequency of breast cancer screening in the UK	Breast cancer (Screening)	Exploring whether trial centres are representative of the other centres in terms of organisational characteristics and adjusting costs for any biases inherent in the organisation of trial centres through the use of sensitivity analysis. Three data sources were used: (1) survey of all breast screening centers within the UK BSP; (2) nationally collected government statistics on outputs for all screening centres; (3) unit costs specially collected from the five screening centres participating in the trial. Trial centres were judged on their representativeness based on medians. One-way and multi-way sensitivity analysis were performed to explore the impact of each characteristic on unit costs for unrepresentative centres	Trial	Trial centres were found to be unrepresentative of non-trial centres for programme size, rural/urban location, number of mobile vans, several assessment procedures and some input levels	Determination of representativeness was not difficult because (1) the median was chosen as a definition of representativeness, (2) many variables were categorical and required subjective judgement and (3) the baseline study was multicentred. Defining an important cost difference is also difficult and a possible solution would be to ask policy-makers. Existence of differences has implications for the selection of centres for inclusion in trials more generally	I

continued

Study	Objective	Clinical area	Methods	Vehicle	Conclusions	Other information	Dimension
Haycox[115]	To test the international transferability of a disease model for UGI disease to different countries	UGI [GORD, ulcer, motility disorders] (PPIs, H2 receptor angonists)	Internationalising a model consists of three steps: (1) identification of variations in national cost-containment policies with their subsequent impact on clinical practice and resource utilisation; (2) measurement of the impact of such variation on health and economic outcomes; (3) valuation of the extent to which such variations altered the global costs and effectiveness for UGI disease. The UK baseline model was adapted to 9 countries (Australia, Austria, Germany, Italy, The Netherlands, Sweden, Switzerland, UK, USA) based on local treatment patterns after discussion with experts. Local unit costs were used to analyse resource consumption within each country. Where certain resources were unavailable or not used, the price of the most commonly used alternative was applied. Estimates concerning the weighted average cost of each category of drug was based upon national market share of each individual drug. Sensitivity analysis was undertaken to evaluate the impact of variations in unit costs or patterns of care	Model	Wide variability exists between countries in the cost of treating patients for UGI disease: (35% GORD, 32.5% ulcer, 32.5% motility disorders): UK = £174.52, Sweden = £312.40, Switzerland = £283.04, Germany = £176.11 A dichotomy exists between the UK and Germany ('the cheap countries') and Sweden and Switzerland ('the expensive countries'). In all major cost categories, the cost of treating patients in the UK and Germany is significantly lower than that in Sweden and Switzerland	The model was sensitive to cost but comparatively insensitive to changes to transition probabilities underlying the model. Much variation can be attributed to differences in the availability and costs of resources between countries. Given the proven transferability of the model, it can be reasonably expected that significant cost advantages underlying implementation in the UK would be equally applicable to other European countries	I

continued

Study	Objective	Clinical area	Methods	Vehicle	Conclusions	Other information	Dimension
Willke[25]	To explore a method for evaluating the generalisability of the overall cost-effectiveness results from a multinational clinical trial to individual countries participating in the trial for individual countries	Subarachnoid haemorrhage (Tirilazad mesylate)	Using regression models that take into account the treatment that a patient receives, a series of exogenous variables (e.g. disease severity), the health outcome of the episode and country-by-treatment and country-by-outcome interaction terms. The model specification allows treatment to affect cost independent of outcome, and allows the outcome to affect the cost of care independent of treatment to enable customisation of cost-effectiveness ratios. Data were used from a Phase 3 trial conducted in nine European countries, Australia and New Zealand of trilazad mesylate for the treatment of aneurysmal subarachnoid haemorrhage. The trial accrued a total of 1023 patients across four different treatment arms (three different doses of tirilazad plus a vehicle-only arm)	Trial	Test of equality of country means $F = 66.6$, $p < 0.0005$ $\chi^2 = 23.6$, $p < 0.0005$. The full treatment effect on costs estimates ranges from a high of \$5845 in additional costs per treated patient in country 2 to a low of \$4812 in cost savings per treated patient in country 5. Differences between country-specific full and net treatment effects on cost are largest in country 1 (\$2230) and lowest in country 2 (\$1303). Cost-effectiveness ratios, overall and by country, were reported. Country-specific utilisation differences, controlling for outcome, clearly contribute to the variation in cost-effectiveness ratios	Cost-decomposition methodology can easily be applied to other studies, including those with non-binary treatment outcomes by using appropriate multivariate techniques and represents at least a partial solution to the common problem of 'customising' multinational trial results for individual countries in the trial. However, it requires that country-specific per patient costs by treatment group are available and will generally not be applicable to countries not contributing cost data to the trial. To help improve economist's ability to use Phase 3 study data, target markets for the clinical or economic end-points of the study should be identified before the study is initiated to ensure that there are a sufficient number of study sites and patients in these markets to develop country-specific estimates. Investigators should develop a data collection strategy that includes resource elements most likely to be affected by practice pattern variation in outcomes expected across markets Non-biological variations in outcomes may be expected across study sites	I

continued

Study	Objective	Clinical area	Methods	Vehicle	Conclusions	Other information	Dimension
Schulman[120]	To describe a framework for the collection and analysis of international resource cost data that can contribute to a consistent and accurate inter-country cost estimation	Subarachnoid haemorrhage (Tirilazad mesylate)	Developing a standardised costing methodology in seven countries by (1) converting total average costs to approximations of marginal unit costs; (2) converting costs from different countries into a common currency using purchasing power parities for the study countries. Where there were still several resource items for which unit costs were unavailable for certain countries, analysts developed an index table, based on a market-basket approach using a method based on physician-work and practice-expense resource-based relative value units. Pairwise comparisons were done for the seven countries where cost data were collected. The basket was defined by six services for which unit cost data were available in all seven countries. The unit cost of each item was multiplied by its mean utilisation in the study to create a weighted total cost for each of the six items for each country and summed to create a market-basket total cost. Where costs are unavailable in all countries, a method based on a set of relative value units developed in the USA was used. The relative value units for physician services were estimated by ascertaining the factors that physicians considered to constitute their work input and other resource costs for medical services and procedures. Separate relative value units were available for the work involved in the procedure (physician-work relative units) and for practice expenses (practice-expense relative value units). By assigning a country-specific cost to the relative value unit, relative value units can be translated into monetary terms. A regression analysis to predict the costs of procedures as a function of the two sets of relative value units and other indicator variables was used to develop an imputation method based on relative value units	Trial	Relative medical cost indices are: Germany (Italy = 0.74, France = 0.56, Sweden = 0.38, UK = 0.42, Australia = 0.49, Spain = 0.52); Italy (Germany = 1.35, France = 0.75, Sweden = 0.51, UK = 0.56, Australia = 0.66, Spain = 0.70); France (Germany = 1.35, Italy = 1.33, Sweden = 0.68, UK = 0.75, Australia = 0.88, Spain = 0.93); Sweden (Germany = 2.64, Italy = 1.96, France = 1.47, UK = 1.10, Australia = 1.30, Spain = 1.37); UK (Germany = 2.41, Italy = 1.79, France = 1.34, Sweden = 0.91, Australia = 1.19, Spain = 1.25); Australia (Germany = 2.03, Italy = 1.51, France = 1.13, Sweden = 0.77, UK = 0.84, Spain = 1.06); Spain (Germany = 1.92, Italy = 1.43, France = 1.07, Sweden = 0.73, UK = 0.80, Australia = 0.95)	If resource-costing exercises are to proceed on a multinational level, researchers must develop resource cost data that are economically comparable. Average total costs are likely to contain overhead or capital costs, which in theory may be invalid for intercountry comparisons	I

continued

Study	Objective	Clinical area	Methods	Vehicle	Conclusions	Other information	Dimension
Cook[131]	To explore the use of statistical methods for testing homogeneity of clinical and economic treatment effects among countries in the Scandinavian Simvastatin Survival Study (4S)	Coronary heart disease (Simvastatin)	Evaluating whether pooling of economic data is appropriate through the conduct of homogeneity tests to detect country-by-treatment interactions. A qualitative or crossover interaction occurs when the treatment effect is positive for patients in some countries and negative in others. A non-crossover interaction occurs when the magnitude, but not the direction, of treatment varies. Resource utilisation data (hospitalisations) were used from the 4S, a controlled trial of simvastatin versus placebo for hypercholesterolaemic patients with existing coronary heart disease in Denmark, Finland, Iceland, Norway and Sweden. Resources were valued using country-specific unit costs	Trial	The results of the test for mortality and hospitalisation rates showed no evidence of a country-by-treatment interaction, indicating that the data could be pooled. Unit cost data were only available for Sweden, so this may have reduced the overall variability in the cost-effectiveness estimates. Although the estimates of the cost-effectiveness ratio for the five countries differed, the Gail and Simon test indicated no evidence of a quantitative interaction. If no treatment-by-country interaction exists, data across countries can be pooled to improve the precision of the estimated cost-effectiveness ratio. Ignoring country and pooling gives a pooled ratio of $59,364 per additional survivor. Alternatively, pooling the data by giving equal weight to each country gives a pooled ratio of $62,653 per additional survivor	Heterogeneity tests for interaction typically have low statistical power. Also, the relative similarity of the five countries in the 4S and the lack of unit costs for all the countries may have led to a bias toward homogeneity that may not be found in other situations	I

continued

Study	Objective	Clinical area	Methods	Vehicle	Conclusions	Other information	Dimension
Coyle[138]	To assess the degree of variation in costs between patients and treatment centres and the determinants of such variation	Head and neck cancer with carcinoma of the bronchus (Conventional radiotherapy, CHART)	Using one-way ANOVA to identify the degree of variation in costs between centres within each subgroup and using OLS regression to assess the degree to which variation could be explained by various factors relating to unit costs, patient characteristics and centre characteristics. Data were used from the CHART, which included 526 patients from 10 UK centres	Trial	The degree of variation within each treatment trial subgroup was similar, but appeared to be greater for CHART therapy than conventional therapy. For both the head and neck and bronchus patients, there was statistically significant variation in costs between centres for CHART patients. For conventional therapy, there was greater variation within centres with no statistically significant variation between centres after Bonferonni correction. For radiotherapy costs, variables identified as determinants of variation were (1) an index of the cost of protocol radiotherapy regimen which is a proxy for overtime payments to staff; (2) a dummy variable denoting whether the treatment centre delivered the CHART treatment during normal working hours. For hospital costs, variables identified as determinants of variation were dummy variables denoting whether the treatment centre used a hostel ward and whether the patient had an advanced tumour stage	The results indicated that significant variation between centres occurred only for CHART and specifically in relation to radiotherapy and hospital costs so the data from individual centres may not be generalisable. Almost all the variation in radiotherapy costs was explained by differences in unit costs between centres primarily due to differences in payment mechanisms for after-hours treatments and the annual use of equipment. For hospital costs, the provision of hostel accommodation and patient characteristics were important determinants of variation between treatments	I

continued

Study	Objective	Clinical area	Methods	Vehicle	Conclusions	Other information	Dimension
Baltussen[114]	To propose a 3-step approach, comprising successive assessment of internal validity, external validity and net impact at the system level to enhance the informative value of economic analysis	Prostatic hyperplasia (BPH) (TUMT versus standard treatment – TURP)	Using a three-step approach: step 1 involves deriving internally valid results through collecting effectiveness parameters and patient-specific data on resource utilisation in an RCT; step 2 involves adapting the results to the real world through taking into account context-specific factors such as the various procedures included in the economic evaluation and specific physician, hospital and healthcare system characteristics; step 3 implies the assessment of costs and outcome changes associated with introducing the new intervention. By additional modelling, cost estimates at the individual level were corrected for the fact that the RCT of TUMT versus standard treatment – TURP – took place in an academic hospital which had larger overhead costs and capacity utilisation. Budget impact was assessed by applying scenario analysis using real-world cost estimates, numbers of patients eligible for TUMT treatment in the future, explicit assumptions regarding demographical, epidemiological and technological developments	Trial	Costs of TUMT and TURP were 28% and 12% lower in the real world than in the RCT. Budget impact analysis encompassed savings of up to $US8 million	Analysis revealed that because of relatively low annual equipment costs, the number of centres through which TUMT is provided affects total costs to the system level only slightly. The number of men seeking BPH treatment following the introduction of TUMT will be a major determinant of the total costs of BPH at the system level	I

continued

Study	Objective	Clinical area	Methods	Vehicle	Conclusions	Other information	Dimension
Goeree[53]	To develop a conceptual framework for selecting hospitals for unit cost estimates in national and international multicentre trials and to test the impact of alternative hospital selection on the cost results	Prelabour rupture of the membranes at term Oxytocin or prostaglandin gel	Consideration for the number of hospitals chosen for unit cost estimates are: (1) the larger the number of hospitals participating in the trial, the less representative or applicable are the costs from one hospital or a small number of hospitals; (2) the method of sampling for selection of hospitals can be systematic or random – the selection of hospitals may be based on key hospital characteristics known to impact on unit costs such as hospital size, level of output, patient mix, teaching status, urban/rural location or extent of staff unionisation; (3) the desired level of subgroup analysis by geographical area – hospital cost variation may be as large within countries as it is between countries. Using PPPs or cross-currency exchange rates are at best only gross-country adjustments. The TermPROM clinical trial investigating the rates of neonatal infection and Caesarian section among 4 management strategies for pregnancies with prelabour rupture of the membranes at term was used which included 72 hospitals across 6 countries with the majority of the patients recruited from Canada, the UK and Australia. Detailed cost information was collected from 6 Canadian hospitals (3 large teaching and 3 small community), 4 UK hospitals (2 large teaching and 2 small community) and 2 Australian hospitals (1 large teaching and 1 small community). Country-specific hospital unit cost estimates were calculated as the midpoint between the highest and lowest cost estimate across hospitals within a country for each resource. Cost analyses were prepared using national and international comparisons	Trial	Basic study results reported no statistically significant differences in the rate of neonatal infection or Caesarean section between any of the 4 management strategies. Across each of the 4 management strategies, the method of selecting and stratifying hospitals resulted in a 30–55% difference between the lowest and highest median estimates. In some cases, the relative ranking of the least and most expensive strategies varied across methods. Furthermore, wide variations were observed in the pairwise statistical comparisons	In both national and international comparisons, both the absolute and relative cost results and statistical comparisons of treatment alternatives varied according to the selection and stratification of hospitals for unit cost estimations. Two solutions are proposed: unit costs should be collected from as many hospitals as possible and a series of multivariate hospital cost studies should be conducted to examine the independent contribution of important cost drivers	I

continued

Study	Objective	Clinical area	Methods	Vehicle	Conclusions	Other information	Dimension
Grieve[84]	To test the validity of the assumption of homogeneity of costs between countries in multinational economic evaluations	Stroke	To illustrate that costs measured in one country may not be generalisable through the collection of country-specific costs. For each patient, durations of hospital stay, use of investigations, average medical time, local prices were collected and directly compared. Pooled data used in multinational economic evaluations may not be applicable to the local setting and instead evaluations should report disaggregated resource use data and county-specific costs which would be more useful to the decision-maker	Trial	The costs of stroke management varied across Europe. The mean length of stay ranged from 12 days in France to 34 days in the UK. The average nursing time per day ranged from 113 minutes in Poland to 318 minutes in Austria. Also unit costs were higher in Western than in Eastern Europe (e.g. a doctor cost $34.21 per hour in France compared with $1.32 in Lithuania). The lowest mean cost per case was in Lithuania ($880, 95% CI 730 to 1030) and the highest in Austria ($8336, 95% CI 6638 to 10033)	Therefore, it can be concluded from this study that hospital costs measured in one European country do not seem generalisable to another and the pooled data from multinational economic evaluations may not be applicable to the local setting	I

continued

Study	Objective	Clinical area	Methods	Vehicle	Conclusions	Other information	Dimension
Koopmanschap[135]	To investigate how costs should be handled in multinational economic evaluation	Breast cancer (Chemotherapy with growth factors)	Using regression analyses to correct for differences in medical practice and calculating the costs of medical consumption as if all patients were treated in one country. Corrected amounts of resource consumption can be multiplied with country-specific costs. The first approach aimed at finding differences in treatment patterns, given relatively homogeneous patient groups. For each of the main categories of medical consumption (e.g. hospital days, consultations, radiotherapy), a statistical analysis tests for country-specific differences in the amount of medical consumption, controlling for variables such as disease stage, treatment arm and age. If statistically 'significant and relevant' differences exist, these should be corrected for. For example, if in country X, 20% fewer laboratory tests are carried out than in country Y (controlling for differences in patient characteristics), patients' laboratory consumption should be corrected downwards when estimating the costs as if all patients were treated in country X. The corrected amounts of medical consumption can then be multiplied with country-specific unit costs. The second approach aimed to explain differences between countries more directly through explaining costs by age, disease state, treatment arm, country and relevant interaction terms	Trial	Approach 1: The range of original hospital costs of €16,900–20,700 narrows to €18,600–20,300 after pooling costs including corrections. The correction factors were modest and were slightly different from the corrections used for pooling because the country-specific corrections (consultations, day-care treatment, laboratory procedures) did not impact upon the largest cost components such as hospitalisation and radiotherapy. Approach 2: The influence of country on costs is not significant for total costs or hospital costs. Adding interaction terms between country and disease state did not produce any significant interaction term and did not improve the explanatory power of the analysis	Two provisional approaches were applied to analyse the differences in medical practice and treatment costs across countries. The first approach explains differences in the amount of medical consumption of separate types of hospital consumption focusing on specific medical services for which a clear homogeneous volume indicator is available. The second approach explains differences in total costs (including drugs) and total hospital costs between countries, but did not detect a significant role for any country parameter, either a simple country dummy or interactions of country and other parameters. This might be as a result of the fact that aggregated hospital costs cannot show differences in costs of specific medical services	I

continued

Study	Objective	Clinical area	Methods	Vehicle	Conclusions	Other information	Dimension
Oostenbrink[159]	To use Markov modelling to convert trial-based cost-effectiveness information from one country to another	Chronic obstructive pulmonary disease Tiotropium, ipratropium, salmeterol, existing therapy	Using Markov modelling to facilitate the transfer of trial-based data to other countries by modelling the difference between treatment groups observed in trials by monthly transition probabilities between disease states and the probabilities of experiencing an event in a given disease state. As resource use is linked to disease states and events, a Markov model can be used to incorporate country-specific resource use and unit costs. Transition probabilities and probabilities of experiencing a severe or non-severe exacerbation were derived from three trial-based cost-effectiveness analyses performed in The Netherlands, Belgium, the USA and 18 other countries. Resource use was segregated into maintenance treatment according to disease severity and treatment associated with exacerbations. Country-specific resource use included the probability and duration of hospitalisation during an exacerbation	Model	The costs of tiotropium were consistently lower than for existing therapy because of the 13–27% lower probability of experiencing a (severe) exacerbation. Costs in the USA were approximately twice as high as in The Netherlands and Belgium. The proportion of total costs due to hospitalisation varied between 30 and 60% depending on the treatment arm and the country	A Markov model in which resource use depends on the disease state and exacerbations, irrespective of treatment group is well suited to convert trial-based economic data on chronic obstructive pulmonary disease to other countries	I
Rutten van-Molken[134]	To compare methods for aggregating resource utilisation and cost data from different countries	Asthma (B2 agonists)	To pool data, two approaches were used: (1) applying country-specific unit prices in national currencies to country-specific volumes of resource use and then converting to US dollars; a regression equation is estimated using dummy variables to control for differences between countries; (2) pooling volumes and adjusting these volumes to better reflect national medical practice and then applying country-specific unit prices in national currencies to the pooled volumes. Data were used from an international 6-month trial of two long-acting bronchodilators to treat patients with asthma	Trial	The choice of aggregation method has considerable impact on estimated costs between the two drugs. Although based on the same underlying volumes, the second method of aggregation demonstrated considerable differences between countries caused by different relative prices of various types of care within countries. However, the second method is more transparent	When there is insufficient power to restrict CEA to one country, the best solution is to pool the volumes of resource use, adjust these pooled volumes to better reflect the national medical practice and apply country-specific unit costs to the adjusted pooled volumes	I

continued

Study	Objective	Clinical area	Methods	Vehicle	Conclusions	Other information	Dimension
Caro[130]	To generalise the results of WOSCOPS to the perspective of any national health service or other organisation responsible for societal costs	Hypercholesterolaemia (Prevastatin)	Deriving a generalised formula expressing the ratio in terms of elements that might be country-specific and those that ought to be more general. It is assumed that the relative treatment effect is generally applicable. However, the baseline risk can be affected by differences in the distribution of risk factors in different populations. With this model, one can calculate the cost-effectiveness for any given baseline risk. The major factor that clearly characterises the country is the cost structure, but full accounting of these details would require separate models. WOSCOPS clinical trial results (undertaken in Scotland) were integrated with Belgian epidemiological and cost data to enable evaluation of the health and economic impact of pravastatin treatment in actual practice in Belgium. Absolute risk reduction = populations reference risk × relative risk reduction. The reference risk in the placebo group was 15.8% over 5 years and use of prevastatin reduced this risk by 22%, resulting in a 3.48% absolute risk reduction. The reference risk was estimated using an exponential regression model that predicted the transition from health to cardiovascular disease without treatment and the relative risk reduction came directly from the trial without adjustment. 744 subjects from the Belgium Interuniversity Research on Nutrition and Health (BIRNH) matched the main WOSCOPS criteria and information on specific values for each of the WOSCOPS risk factors was obtained, and the individual patient's probability of a cardiovascular event was estimated using a regression equation. The local reference risk and relative risk reduction were applied in a Markov model analysis that ran for a total period of five years. The risk determines the number of individuals who will manifest the disease for the first time during that month and is estimated as the instantaneous transition rate (hazard) for both treatment arms from the cumulative incidence (risk). Costs were assigned to events occurring	Trial Model	Absolute relative risk reduction in Belgium parallels that observed in WOSCOPS (3.19% vs 3.48%). Use of prevastatin would prevent 297 out of 10,000 from developing overt cardiovascular disease. Cost-effectiveness ratio is €29,800 per life year gained (benefits discounted). The corresponding ratios for the UK is €31,400	Clinical and economic findings from WOSCOPS are generalisable to other populations with similar characteristics in terms of gender, age and cholesterol level	I

continued

Study	Objective	Clinical area	Methods	Vehicle	Conclusions	Other information	Dimension
Rutten-van Molken[133]	To determine the relative economic consequences of treating asthmatic subjects with twice-daily dry powder formoterol 12 μg as compared with salmeterol 50 μg after correction for country differences	Asthma (B2 agonists)	Using OLS regression analyses to correct for possible differences between countries and patients when testing the impact of the trial medication on total costs. Independent variables included age, duration of illness in years, average peak flow during run-in, average day- and night-time symptom score during run-in and dummy variables for the trial drug, country, gender, smoking and having a paid job. Differences in definitions, costing methodology and uncertainty associated with data sources explain part of the variation in unit prices so it is difficult to pool the cost data from the 6 participating countries. Therefore, the importance of separating resource use from costs rather than simply reporting and analysing total costs is emphasised	Trial (randomised open label)	OLS regression analyses of the pooled log transformed costs confirmed that there were no statistically significant differences in costs between formoterol and salmeterol. Proportion of variation in costs explained in the regression model was 24% for total costs and 20% for direct costs. Significant differences exist between countries. Swiss patients had significantly higher costs than patients from other countries and Italian patients had lower costs	The regression analysis corrected for possible differences between countries and patients when testing the impact of the trial medication on total costs. Differences in definitions, costing methodology and uncertainty associated with data sources explain part of the variation in unit prices. It is difficult to pool the cost data from the 6 participating countries. Therefore, the authors emphasise the importance of separating resource use from costs rather than simply reporting and analysing total costs	I

continued

Study	Objective	Clinical area	Methods	Vehicle	Conclusions	Other information	Dimension
Menzin[116]	To explore how data from pivotal clinical trials in one setting can serve as a basis for economic evaluations in others using the example of recombinant human deoxyribonuclease (rhDNase) to improve pulmonary function in patients with cystic fibrosis	RTIs (rhDNase)	Modelling based on the assumption that the relative risk reduction is generalisable to other settings, but resource utilisation is country-specific. Information is ascertained through expert opinion on 'practice pattern' parameters believed likely to vary across countries such as the likelihood of hospitalisation and the associated mean length of stay in hospital. A decision model was used as a means of extrapolating from the US trial data to the other locations. The US Phase 3 trial compared two different doses of rhDNase with vehicle (placebo). In the trial, patients were treated for 24 weeks and the outcome measures included change in pulmonary function (FEV1) and incidence of RTIs requiring parenteral therapy. In terms of economic endpoints, data on hospital admissions, inpatient days and days of oral and intravenous antibiotic therapy were collected. The US rate of hospitalisation and length of stay data were substituted by the relevant country-specific estimates for France, Germany, Italy and the UK	Trial Model	US trial demonstrated a significant reduction in RTI-related hospital admissions with rhDNase (0.41 vs 0.56 for placebo; $p < 0.05$) and in days of RTI-related outpatient intravenous antibiotic therapy (2.9 vs 4.4; $p < 0.05$). Compared with placebo, the cost of treating RTIs over 24 weeks was $1682 less among patients receiving rhDNase once daily, primarily owing to reductions in the cost of hospitalisation	Pricing US resource use observed in the trial in local currency demonstrated savings from £434 ($711) in the UK to FF 7010 ($1064) in France. Since US trial-based resource estimates were potentially misleading, adjustments were made based on the variability in two key parameters (rate of hospitalisation and length of stay). Overall effect was to reduce slightly the estimates of savings in Italy (from $908 to $607) and in France (from $1064 to $850). In Germany, there was a very small increase in savings from treating RTI	I

continued

Study	Objective	Clinical area	Methods	Vehicle	Conclusions	Other information	Dimension
Shih[142]	To overcome some of the bias inherent in RCT data while avoiding some of the common pitfalls associated with the use of observational data	End-stage renal disease (Erthyropoietin, blood transfusion)	Using a decomposition technique between different population subgroups based on the principle of removing population effects from observed outcome effects and identification of the separate components of the outcome effect (treatment and population) is key. The treatment effect is the difference in relative risk applied to the treatment group, whereas the population effect is an adjustment for differences in 2 study populations. Multivariate regression models are used to regress costs on sets of risk factors such as age, gender, race, disease group and logistic/linear regression used to estimate branch probabilities and payoffs within a decision-modelling framework. The costs of 2 treatments for anaemia of renal failure (erthyropoietin, blood transfusion) in an outpatient setting were modelled	Model	Under standard methods of decision analysis, an increase of US$7032 per patient following erythropoietin coverage was observed. Using the decomposition technique, this was US$6172, the difference coming from the population effect	Failure to remove population effects from observed outcome effects could lead to biased decision-making. Although not directly observable, the population effect can be imputed from secondary data. The decomposition and imputing technique allows for a more meaningful interpretation of the results for the purpose of policy analysis	I
Glick[121]	To investigate for collection of hospital unit cost data for use in multinational clinical trials, for how many types of hospitalisation and how many countries should estimates be obtained	Not specified	To investigate whether reliable methods for imputing hospitalisation costs exist to allow one to economise on data collection, subsets of hospitalisation types were randomly sampled and used to develop imputation regressions. It was assumed that the unit cost estimates for 47 types of hospitalisations collected in 4 European countries represented the universe of hospitalisations which was used as a population estimate against which error associated with imputations in population samples could be measured. To determine how many types of hospitalisation estimates should be obtained, (1) subsets were randomly sampled and used to develop imputation regressions; (2) these regression results were used to impute costs; and (3) measures of imputation error within each sample were estimated. To determine how many countries, a similar analysis was performed but with countries	Trial	The imputation error decreased as the number of types of hospitalisations and countries sampled increased, but the rate of reduction in error shrank. Further, the error was minimised by obtaining estimates for fewer types of hospitalisations from more countries	The availability of reliable methods for imputing hospitalisation costs allows one to economise on data collection. Collecting a small number of estimates (in this study, ~25) in as many countries as is feasible minimises imputation error	I

continued

Study	Objective	Clinical area	Methods	Vehicle	Conclusions	Other information	Dimension
Spath[147]	To propose a three-step approach to evaluate the potential of economic studies for transfer	Breast cancer (Adjuvant therapy)	Using a checklist approach to identify relevant economic evaluations based on four inclusion criteria: perspective, number of competing options, appropriate therapies and relevance of therapies to the healthcare system concerned. This is followed by an analysis of the eligibility for transfer based on five indicators based on three dimensions: settings in which the studies might be used, the transferability of health outcome data and the transferability of resource utilisation data. Economic evaluations were identified through electronic databases and searches of bibliographies and independently reviewed by a health economist and an oncologist	Review	Of the 26 studies identified in the literature, 20 did not satisfy four inclusion criteria: no identification of perspective, no comparison of two or more therapies, no comprehensive description of the competing therapies or the therapies were not used within the French healthcare system. Six studies were appraised but none of these studies were considered relevant to the French healthcare system	The main reason for this conclusion was that the cost data were not reported in a transparent way. The authors suggested that if the studies had been suitable, the next steps for analysing the transferability of published economic evaluations would have been to check whether the data applied to their local setting and, if not, to assess whether it should be adapted	1

1, Location; 2, time.
BCR, benefit-to-cost ratio; BPH, prostatic hyperplasia; CHART, continuous hyperfractionated accelerated radiotherapy; PPI, proton-pump inhibitor; RTI, respiratory tract infections; TUMT, transurethral microwave therapy; TURP, transurethral resection of the prostate.

Appendix 3

List of papers included in the systematic review of economic evaluations alongside multilocation trials in Chapter 5

A cost–benefit analysis using a willingness-to-pay questionnaire of intranasal budesonide for seasonal allergic rhinitis. *Ann Allergy Asthma Immunol* 2000; **84**:55–62.

Armstrong SH, Ruckley CV. Use of a fibrous dressing in exuding leg ulcers. *J Wound Care* 1997;**6**;322–4.

Bailey TC, Ritchie DJ, McMullin ST, Kahn M, Reichley RM, Casabar E, *et al*. A randomized, prospective evaluation of an interventional program to discontinue intravenous antibiotics at two tertiary care teaching institutions. *Pharmacotherapy* 1997;**17**:277–81.

Bassan R, Lerede T, Di Bona E, Rossi G, Pogliani E, Rambaldi A, *et al*. Granulocyte colony-stimulating factor (G-CSF, Filgrastim) after or during an intensive remission induction therapy for adult acute lymphoblastic leukaemia: effects, role of patient pre-treatment characteristics, and costs. *Leuk Lymphoma* 1997;**26**:153–61.

Bennett-Guerrero E, Sorohan JG, Gurevich ML, Kazanjian PE, Levy RR, Barbera AV, *et al*. Cost–benefit and efficacy of aprotinin compared with epsilon-aminocaproic acid in patients having repeated cardiac operations:a randomized, blinded clinical trial. *Anesthesiology* 1997;**87**:1373–80.

van Bergen PFMM, Jonker JJC, van Hot BA, van Domburg RT, Deckers JW, Azar AJ, *et al*. Costs and effects of long-term oral anticoagulant treatment after myocardial infarction. *JAMA* 1995;**273**:925–8.

Boccalon H, Elias A, Chale J-J, Cadene A, Gabriel S, for the Vascular Midi-Pyrenees Network Group. Clinical outcome and cost of hospital vs home treatment of proximal deep vein thrombosis with a low-molecular-weight heparin: the Vascular Midi-Pyrenees Study. *Arch Intern Med* 2000;**160**:1769–73.

Booth PC, Wells NEJ, Morrison AK. A comparison of the cost effectiveness of alternative prophylactic therapies in childhood asthma. *Pharmacoeconomics* 1996; **10**:262–8.

Burton LC, Steinwachs DM, German PS, Shapiro S, Brant LJ, Richards TM, *et al*. Preventive services for the elderly: would coverage affect utilisation and costs under Medicare? *Am J Public Health* 1995;**85**:387–91.

Bytzer P, Moller Hansen JM, Schaffalitzky de Muckadell OB. Empirical H2-blocker therapy or prompt endoscopy in management of dyspepsia. *Lancet* 1994;**343**:811–16.

Canard JM, Cellier C, Houcke P, Laurent J, Gorce D, Landi B. Prospective multicentre study comparing a standard reusable sphincterotome with a disposable triple-lumen sphincterotome. *Gastrointest Endosc* 2000; **51**:704–7.

Choiniere M, Rittenhouse BE, Perreault S, Chartrand D, Rousseau P, Smith B, *et al*. Efficacy and costs of patient-controlled analgesia versus regularly administered intramuscular opioid therapy. *Anesthesiology* 1998; **89**:1377–88.

Cunningham D, Gore M, Davidson N, Miocevich M, Manchanda M, Wells N. The real costs of emesis: an economic analysis of ondansetron vs metoclopramide in controlling emesis in patients receiving chemotherapy for cancer. *Eur J Cancer Part A* 1993;**29**:303–6.

Dasbach EJ, Rich MW, Segal R, Gerth WC, Carides GW, Cook JR, *et al*. The cost-effectiveness of losartan versus captopril in patients with symptomatic heart failure. *Cardiology* 1999;**91**:189–94.

Davey P, Craig AM, Hau C, Malek M. Cost-effectiveness of prophylactic nasal mupirocin in patients undergoing peritoneal dialysis based on a randomized, placebo-controlled trial. *J Antimicrob Chemother* 1999;**43**;105–12.

Detournay B, Huet X, Fognani F, Montalescot G. Economic evaluation of enoxaparin sodium versus heparin in unstable angina: a French sub-study of the ESSENCE trial. *Pharmacoeconomics* 2000;**18**:83–9.

van Dijk LC, Seerden R, van Urk H, Wittens CHA. Comparison of cost affecting parameters and costs of the "closed" and "open" in situ bypass technique. *Eur J Vasc Endovasc Surg* 1997;**13**:460–3.

Drummond N, Abdall M, Buckingham JK, Beattie JAG, Lindsay T, Osman LM, *et al*. Grampian Asthma Study of Integrated Care (GRASSIC). Integrated care for asthma – a clinical, social, and economic evaluation. *BMJ* 1994; **308**:559–64.

Ellis E, Koblin W, Irvine M J, Legare J, Logan AG. Small, blue collar work site hypertension screening: a cost-effectiveness study. *J Occup Med* 1994;**36**:346–55.

Ellis SL, Carter BL, Malone DC, Billups SJ, Okano GJ, Valuck RJ, *et al*. Clinical and economic impact of ambulatory care clinical pharmacists in management of dyslipidemia in older adults: the IMPROVE study. *Pharmacotherapy* 2000;**20**:1508–16.

Erhardt L, Ball S, Andersson F, Bergentoft P, Martinez C. Cost effectiveness in the treatment of heart failure with ramipril: a Swedish substudy of the AIRE study [published erratum appears in *Pharmacoeconomics* 1997;**12**:706]. *Pharmacoeconomics* 1997;**12**:256–66.

Fagnani F, Bouvenot G, Valat J-P, Bardin T, Berdah L, Lafuma A, *et al*. Medico-economic analysis of diacerein with or without standard therapy in the treatment of osteoarthritis. *Pharmacoeconomics* 1998;**13**:135–46.

Franz MJ, Splett PL, Monk A, Barry B, McClain K, Weaver T, *et al*. Cost-effectiveness of medical nutrition therapy provided by dietitians for persons with non-insulin-dependent diabetes mellitus. *J Am Diet Assoc* 1995;**95**:1018–24.

Friedman M, Serby CW, Menjoge SS, Wilson JD, Hilleman DE, Witek TJ. Pharmacoeconomic evaluation of a combination of ipratropium plus albuterol compared with ipratropium alone and albuterol alone in COPD. *Chest* 1999;**115**:635–41.

Gafni A, Goeree R, Myhr TL, Hannah ME, Blackhouse G, Willan AR, *et al*. Induction of labour versus expectant management for prelabour rupture of the membranes at term: an economic evaluation. *Can Med Assoc J* 1997;**157**:1519–25.

Glick H, Willke R, Polsky D, Llana T, Alves WM, Kassell N, Schulman K. Economic analysis of tirilazad mesylate for aneurysmal subarachnoid hemorrhage: economic evaluation of a phase III clinical trial in Europe and Australia. *Int J Technol Assess Health Care* 1998;**14**:145–60.

Gournay K, Brooking J. The community psychiatric nurse in primary care: an economic analysis. *J Adv Nurs* 1995;**22**:769–78.

Gray A, Raikou M, McGuire A, Fenn P, Stevens R, Cull C, *et al*., on behalf of the United Kingdom Prospective Diabetes Study Group. Cost effectiveness of an intensive blood glucose control policy in patients with type 2 diabetes: economic analysis alongside randomised controlled trial (UKPDS 41). *BMJ* 2000; **320**:1373–8.

Griffiths R I, Anderson G F, Powe N R, Oliveras E, Herbert R J, Grant CC, *et al*. Economic impact of immunization against rotavirus gastroenteritis: evidence from a clinical trial. *Arch Pediatr Adolesc Med* 1995; **149**:407–14.

Grines CL, Marsalese DL, Brodie B, Griffin J, Donohue B, Constantini CR, *et al*. Safety and cost-effectiveness of early discharge after primary angioplasty in low risk patients with acute myocardial infarction. *J Am Coll Cardiol* 1998;**31**:967–72.

Grossman R, Mukherjee J, Vaughan D, Eastwood C, Cook R, LaForge J, *et al*. A 1-year community-based health economic study of ciprofloxacin vs usual antibiotic treatment in acute exacerbations of chronic bronchitis: the Canadian Ciprofloxacin Health Economic Study Group. *Chest* 1998;**113**:131–41.

Grunfeld E, Gray A, Mant D, Yudkin P, Adewuyi-Dalton R, Coyle D, *et al*. Follow up of breast cancer in primary care vs specialist care: results of an economic evaluation. *Br J Cancer* 1999;**79**:1227–33.

Guthrie E, Moorey J, Margison F, Barker H, Palmer S, McGrath G, *et al*. Cost-effectiveness of brief psychodynamic-interpersonal therapy in high utilizers of psychiatric services. *Arch Gen Psychiatry* 1999; **56**:519–26.

Hansson C. The effects of cadexomer iodine paste in the treatment of venous leg ulcers compared with hydrocolloid dressing and paraffin gauze dressing. *Int J Dermatol* 1998;**37**:390–6.

Harrington R, Peters S, Green J, Byford S, Woods J, McGowan R. Randomised comparison of the effectiveness and costs of community and hospital based mental health services for children with behavioural disorders. *BMJ* 2000;**321**:1–5.

Harvey I, Nelson SJ, Lyons RA, Unwin C, Monaghan S, Peters T. A randomized controlled trial and economic evaluation of counselling in primary care. *Br J Gen Pract* 1998;**48**:1043–8.

Hamilton SH, Revicki DA, Edgell ET, Genduso LA, Tollefson G. Clinical and economic outcomes of olanzapine compared with haloperidol for schizophrenia: results from a randomised clinical trial. *Pharmacoeconomics* 1999;**15**:469–80.

Hillman AL, Schwartz JS, Willian MK, Peskin E, Roehrborn CG, Oesterling JE, *et al*. The cost-effectiveness of terazosin and placebo in the treatment of moderate to severe benign prostatic hyperplasia. *Urology* 1996;**47**:169–78.

Houcke P, Maffioli C, Corallo J, Bouxinsauzet A, Martin AL. Comparison of Rocgel® with an h-2-antihistamine on the symptomatology of gastroesophageal reflux without esophagitis. *Ann Gastroenterol Hepatol (Paris)* 1995;**31**:264–7 (in French).

Hull RD, Raskob GE, Rosenbloom D, Pineo GF, Lerner RG, Gafni A, *et al*. Treatment of proximal vein thrombosis with subcutaneous low-molecular-weight heparin vs intravenous heparin. *Arch Intern Med* 1997;**157**:289–94.

Hutchins SS, Rosenthal J, Eason P, Swint E, Guerrero H, Hadler S. Effectiveness and cost-effectiveness of linking the special supplemental program for women, infants, and children (WIC) and immunization activities. *J Public Health Policy* 1999;**20**:408–26.

Jones R, Tait C, Sladen G, Weston-Baker J. A trial of a test-and-treat strategy for helicobacter pylori positive dyspeptic patients in general practice. *Int J Clin Pract* 1999;**53**:413–16.

Jonsson B, Cook JR, Pedersen TR. The cost-effectiveness of lipid lowering in patients with diabetes results from the 4S trial. *Diabetologia* 1999;**42**:1293–301.

Jonsson B, Johannesson M, Kjekshus J, Olsson AG, Pedersen TR, Wedel H. Cost-effectiveness of cholesterol

lowering: results from the Scandinavian Simvastatin Survival Study (4S). *Eur Heart J* 1996;**17**:1001–7.

Kald A, Anderberg B, Carlsson P, Park PO, Smedh K. Surgical outcome and cost-minimisation-analyses of laparoscopic and open hernia repair: a randomised prospective trial with one year follow up. *Eur J Surg* 1997;**163**:505–10.

Karalus NC, Garrett JE, Lang SDR, Scott WG, Leng RA, Kostalas GN, *et al*. A clinical and economic comparison of roxithromycin 150 mg twice daily vs amoxicillin 500 mg/calvulanic acid 125 mg three times daily for the treatment of lower respiratory tract infections in general practice. *Drug Invest* 1994;**8**:179–90.

Kumpusalo E, Neittaanmaki L, Halonen P, Pekkarinen H, Penttila I, Parviainen M. Finnish healthy village study: impact and outcomes of a low-cost local health promotion programme. *Health Promot Int* 1996; **11**:105–15.

Laloux P, Vakaet A, Monseu G, Jacquy J, Bourgeois P, van der Linden C. Subcutaneous sumatriptan compared with usual acute treatments for migraine: clinical and pharmacoeconomic evaluation. *Acta Neurol Belg* 1998;**98**:332–41.

Lawrence K, McWhinnie D, Goodwin A, Gray A, Gordon J, Storie J, *et al*. An economic evaluation of laparoscopic versus open inguinal hernia repair. *J Public Health Med* 1996;**18**:41–8.

LePen C, Lilliu H, Keller T, Fiessinger S. The economics of TRACE: a cost-effectiveness analysis of trandolapril in postinfarction patients with left ventricular dysfunction. *Pharmacoeconomics* 1998;**14**:49–58.

Liem MSL, Halsema JAM, Van der Graaf Y, Schrijvers AJP, van Vroonhoven TJMV. Cost-effectiveness of extraperitoneal laparoscopic inguinal hernia repair: a randomized comparison with conventional herniorrhaphy. *Ann Surg* 1997;**226**:668–75.

Liljas B, Stahl E, Pauwels RA. Cost effectiveness analysis of a dry powder inhaler (Turbuhaler) versus a pressurised metered dose inhaler in patients with asthma. *Pharmacoeconomics* 1997;**12**:267–77.

Lundback B, Jenkins C, Price MJ, Thwaites RMA. Cost-effectiveness of salmeterol/fluticasone propionate combination product 50/250 μg twice daily and budesonide 800 μg twice daily in the treatment of adults and adolescents with asthma. *Respir Med* 2000; **94**:724–32

Malmivaara A, Hakkinen U, Aro T, Heinrichs ML, Koskenniemi L, Kuosma E, *et al*. The treatment of acute low back pain: bed rest, exercises, or ordinary activity? *N Engl J Med* 1995;**332**:351–5.

Mark DB, Cowper PA, Berkowitz SD, Davidson-Ray L, DeLong ER, Turpie AGG, *et al*. Economic assessment of low-molecular-weight heparin (enoxaparin) versus unfractionated heparin in acute coronary syndrome patients: results from the ESSENCE randomized trial. *Circulation* 1998:**97**:1702–7.

Mark DB, Harrington RA, Lincoff AM, Califf RM, Nelson CL, Tsiatis AA, *et al*. Cost-effectiveness of platelet glycoprotein IIb/IIIa inhibition with eptifibatide in patients with non-ST-elevation acute coronary syndromes. *Circulation* 2000;**101**:366–71.

Martinez C, Ball SG. Cost-effectiveness of ramipril therapy for patients with clinical evidence of heart failure after acute myocardial infarction. *Br J Clin Pract* 1995;(Suppl 78):26–32.

Meenan RT, Stevens VJ, Hornbrook M C, La Chance P-A, Glasgow RE, Hollis JF, *et al*. Cost-effectiveness of a hospital-based smoking cessation intervention. *Med Care* 1998;**36**:670–8.

Mol BWJ, Hajenius PJ, Engelsbel S, Ankum WM, Hemrika DJ, Van der Veen F, *et al*. Treatment of tubal pregnancy in The Netherlands: an economic comparison of systemic methotrexate administration and laparoscopic salpingostomy. *Am J Obstet Gynecol* 1999;**181**:945–51.

Monteon F, Correa-Rotter R, Paniagua R, Amato D, Hurtado ME, Medina JL, *et al*. Prevention of peritonitis with disconnect systems in CAPD: a randomised controlled trial. *Kidney Int* 1998;**54**:2123–8.

Morrell CJ, Walters SJ, Dixon S, Collins KA, Brereton LML, Peters J, *et al*. Cost effectiveness of community leg ulcer clinics: randomised controlled trial. *BMJ* 1998;**316**;1487–91.

Murphy MA, Maze NM, Boyd JL, Quick DC, Buss DD. Cost–benefit comparison: holmium laser versus electrocautery in arthroscopic acromioplasty. *J Shoulder Elbow Surg* 1999;**8**:275–8.

Naglak M, Mitchell DC, Kris-Etherton P, Harkness W, Pearson TA. What to consider when conducting a cost-effectiveness analysis in a clinical setting. *J Am Diet Assoc* 1998;**98**:1149–54.

Naylor MD, Brooten D, Campbell R, Jacobsen R, Mezey MD, Pauly MV, *et al*. Comprehensive discharge planning and home follow-up of hospitalised elders: a randomized clinical trial. *JAMA* 1999;**281**:613–20.

Nicolle LE, Kirshen A, Boustcha E, Montgomery P. Treatment of moderate-to-severe pneumonia in elderly, long-term care facility residents. *Infect Dis Clin Pract* 1996;**5**:130–6.

Nord E, Wisloff F, Hjorth M, Westin J. Cost–utility analysis of melphalan plus prednisone with or without interferon-Alpha2b in newly diagnosed multiple myeloma: results from a randomised controlled trial. *Pharmacoeconomics* 1997;**12**:89–103.

O'Brien B, Levine M, Willan A, Goeree R, Haley S, Blackhouse G, *et al*. Economic evaluation of outpatient treatment with low-molecular-weight heparin for proximal vein thrombosis. *Arch Intern Med* 1999; **159**:2298–304.

O'Byrne P, Cuddy L, Taylor DW, Birch S, Morris J, Syrotuik J. Efficacy and cost benefit of inhaled

corticosteroids in patients considered to have mild asthma in primary care practice. *Can Respir J* 1996; **3**:169–75.

Oster G, Huse DM, Lacey MJ, Regan MM, Fuchs HJ. Effects of recombinant human dnase therapy on healthcare use and costs in patients with cystic fibrosis. *Ann Pharmacother*. 1995;**29**:459–64.

Paganini AM, Lezoche E, Carle F, Carlei F, Favretti F, Feliciotti F, *et al*. A randomized, controlled, clinical study of laparoscopic vs open tension-free inguinal hernia repair. *Surg Endosc* 1998;**12**:979–86.

Pedersen TR, Kjekshus J, Berg K, Olsson AG, Wilhelmsen L, Wedel H, *et al*. Cholesterol lowering and the use of healthcare resources: results of the Scandinavian simvastatin survival study. *Circulation* 1996; **93**:1796–802.

Pelc A, Portier H, Gehanno P, Fiessinger S, Ichou F. Cost saving of 5-day therapy with cefpodoxime proxetil versus standard 10-day beta-lactam therapy for recurrent pharyngotonsillitis in adults. *Pharmacoeconomics* 1996;**10**:239–50.

Petrou S, Coyle D, Fraser W D. Cost-effectiveness of a delayed pushing policy for patients with epidural anesthesia. *Am J Obstet Gynecol* 2000;**182**:1158–64.

Poynard T, Ottignon Y, Pauphilet C, Agostini H. A randomized trial of the usefulness of endoscopy after an initial treatment of non-severe esophagitis. *Gastroenterol Clin Biol*. 1997;**21**:497–502.

Poynard T, Vernisse B, Agostini H. Randomized, multicentre comparison of sodium alginate and cisapride in the symptomatic treatment of uncomplicated gastro-oesophageal reflux. *Aliment Pharmacol Ther* 1998;**12**:159–65.

Rabah M, Mason D, Muller DWM, Hundley R, Kugelmass AD, Weiner B, *et al*. Heparin after percutaneous intervention (HAPI): a prospective multicenter randomized trial of three heparin regimens after successful coronary intervention. *J Am Coll Cardiol* 1999;**34**:461–7.

Rajan R, Gafni A, Levine M, Hirsh J, Gent M. Very low-dose warfarin prophylaxis to prevent thromboembolism in women with metastatic breast cancer receiving chemotherapy: an economic evaluation. *J Clin Oncol* 1995;**13**:42–6.

Rodriguez A, Ayala F, Bernardi V, Santaera O, Marchand E, Pardinas C, *et al*. Optimal coronary balloon angioplasty with provisional stenting versus primary stent (OCBAS): immediate and long-term follow-up results. *J Am Coll Cardiol* 1998;**32**:1351–7.

Roseveare CD, Patel P, Simmonds N, Goggin PM, Kimble J, Shepherd HA. Metal stents improve dysphagia, nutrition and survival in malignant oesophageal stenosis:a randomized controlled trial comparing modified Gianturco Z-stents with plastic Atkinson tubes. *Eur J Gastroenterol Hepatol* 1998; **10**:653–7.

Rutten-van Molken MPMH, van Doorslaer EKA, Till MD. Cost-effectiveness analysis of formoterol versus salmeterol in patients with asthma. *Pharmacoeconomics* 1998;**14**:671–84.

Sacristan JA, Kennedy-Martin T, Rosell R, Cardenal F, Anton A, Lomas M, *et al*. Economic evaluation in a randomized phase III clinical trial comparing gemcitabine/cisplatin and etoposide/cisplatin in non-small cell lung cancer. *Lung Cancer* 2000;**28**:97–107.

Sanchez Garcia M, Cambronero Galache JA, Lopez Diaz J, Cerda Cerda E, Rubio Blasco J, Gomez Aguinaga MA, *et al*. Effectiveness and cost of selective decontamination of the digestive tract in critically ill intubated patients: a randomized, double-blind, placebo-controlled, multicenter trial. *Am J Respir Crit Care Med* 1998;**158**:908–16.

Serruys PW, van Hout B, Bonnier H, Legrand V, Eulogio Garcia, Macaya C, *et al*., for the Benestent Study Group. Randomised comparison of implantation of heparin-coated stents with balloon angioplasty in selected patients with coronary artery disease (Benestent II). *Lancet* 1998;**352**: 673–81.

Severens JL, Prahl C, Kuijpers-Jagtman AM, Prahl-Andersen B. Short-term cost-effectiveness analysis of presurgical orthopedic treatment in children with complete unilateral cleft lip and palate. *Cleft Palate Craniofac J* 1998;**35**:222–6.

Severson HH, Andrews JA, Lichtenstein E, Wall M, Akers L. Reducing maternal smoking and relapse: long-term evaluation of a pediatric intervention. *Prev Med* 1997;**26**:120–30.

Schmidt-Matthiesen A, Roding H, Windolf J, Sommerfeldt D, Gutt C, Pannike A, *et al*. A prospective, randomised comparison of single- vs. multiple-dose antibiotic prophylaxis in penetrating trauma. *Chemotherapy* 1999;**45**:380–91.

Schnitzler MA, Woodward RS, Lowell JA, Singer GG, Brennan DC. High risk kidney transplant rejection treatment: cost savings from thymoglobulin. *Transplant Proc* 1999;**31**:269–71.

Schulman KA, Buxton M, Glick H, Sculpher M, Guzman G, Kong J, *et al*. Results of the economic evaluation of the First Study: a multinational prospective economic evaluation. *Int J Technol Assess Health Care* 1996;**12**:698–713.

Schulman KA, Dorsainvil D, Yabroff KR, DiCesare J, Felser J, Eisenberg JM, *et al*. Prospective economic evaluation accompanying a trial of GM-CSF/IL-3 in patients undergoing autologous bone marrow transplantation for Hodgkins and non-Hodgkins lymphoma. *Bone Marrow Transplant* 1998;**21**:607–14.

Scott WG, Cooper B C, Scott HM. Pharmacoeconomic evaluation of roxithromycin versus amoxycillin clavulanic acid in a community-acquired lower respiratory tract infection study. *Infection* 1995; **23** (Suppl 1):S21–4.

Sculpher MJ, Seed P, Henderson RA, Buxton MJ, Pocock SJ, Parker J, *et al*. Health service costs of coronary angioplasty and coronary artery bypass surgery: the Randomised Intervention Treatment of Angina (RITA) trial. *Lancet* 1994;**344**:927–33.

Skargren EI, Oberg BE, Carlsson PG, Gade M. Cost and effectiveness analysis of chiropractic and physiotherapy treatment for low back and neck pain: six-month follow-up. *Spine* 1997;**22**:2167–77.

Stone GW, Grines CL, Rothbaum D, Browne KF, O'Keefe J, Overlie PA, *et al*. Analysis of the relative costs and effectiveness of primary angioplasty versus tissue-type plasminogen activator:the primary angioplasty in myocardial infarction (PAMI) trial. *J Am Coll Cardiol* 1997;**29**:901–7.

Suver J, Arikian SR, Doyle JJ, Sweeney SW, Hagan M. Use of anesthesia selection in controlling surgery costs in an HMO Hospital. *Clin Ther* 1995;**17**:561–71.

Sweat M, Gregorich S, Sangiwa G, Furlonge C, Balmer D, Kamenga C, *et al*. Cost-effectiveness of voluntary HIV-1 counselling and testing in reducing sexual transmission of HIV-1 in Kenya and Tanzania. *Lancet* 2000;**356**:113–21.

The Canadian Lung Oncology Group. Investigation for mediastinal disease in patients with apparently operable lung cancer. *Ann Thorac Surg* 1995;**60**:1382–9.

Torrance G, Walker V, Grossman R, Mukherjee J, Vaughan D, La Forge J, *et al*. Economic evaluation of ciprofloxacin compared with usual antibacterial care for the treatment of acute exacerbations of chronic bronchitis in patients followed for 1 year. *Pharmacoeconomics* 1999;**16**:499–520.

Tsevat J, Duke D, Goldman L, Pfeffer M A, Lamas GA, Soukup JR, *et al*. Cost-effectiveness of captopril therapy after myocardial infarction. *J Am Coll Cardiol* 1995; **26**:914–19.

Venning P, Durie A, Roland M, Roberts C, Leese B. Randomised controlled trial comparing cost effectiveness of general practitioners and nurse practitioners in primary care. *BMJ* 2000;**320**:1048–53.

Vergnenegre A, Perol M, Pham E, Balmes P, Blanchon F, Bonnaud F, *et al*. Cost analysis of hospital treatment: two chemotherapic regimens for non-surgical non-small cell lung cancer. *Lung Cancer* 1996;**14**:31–44.

Webb A, Cunningham D, Scarffe J H, Harper P, Norman A, Joffe J K, *et al*. Randomised trial comparing epirubicin, cisplatin, and fluorouracil versus fluorouracil, doxorubicin, and methotrexate in advanced esophagogastric cancer. *J Clin Oncol* 1997; **15**:261–7.

Weaver WD, Reisman MA, Griffin JJ, Buller CE, Leimgruber PP, Henry T, *et al*., for the OPUS-1 Investigators. Optimum percutaneous transluminal coronary angioplasty compared with routine stent strategy trial (OPUS-1): a randomised trial. *Lancet* 2000; **355**:2199–203.

Weintraub WS, Culler S, Boccuzzi SJ, Cook JR, Kosinski AS, Cohen DJ, *et al*. Economic impact of GPIIB/IIIA blockade after high-risk angioplasty: results from the RESTORE trial. *J Am Coll Cardiol* 1999; **34**:1061–6.

Wever EFD, Hauer RNW, Schrijvers G, van Capelle FJL, Tijssen JGP, Crijns HJGM, *et al*. Cost-effectiveness of implantable defibrillator as first-choice therapy versus electrophysiologically guided, tiered strategy in postinfarct sudden death survivors: a randomized study. *Circulation* 1996;**93**:489–96.

White DJ, Billingham C, Chapman S, Drake S, Jayaweera D, Jones S, *et al*. Podophyllin 05% or 20% v podophyllotoxin 05% for the self treatment of penile warts: a double blind randomised study. *Genitourin Med* 1997;**73**:184–7.

Wolter JM, Bowler SD, Nolan PJ, McCormack JG. Home intravenous therapy in cystic fibrosis: a prospective randomised trial examining clinical, quality of life and cost aspects. *Eur Respir J* 1997;**10**:896–900.

Wootton R, Bloomer SE, Corbett R, Eedy DJ, Hicks N, Lotery HE, *et al*. Multicentre randomised control trial comparing real time teledermatology with conventional outpatient dermatological care: societal cost–benefit analysis. *BMJ* 2000;**320**:1252–6.

Appendix 4

Search strategies employed for electronic databases relating to the review detailed in Chapter 7

MEDLINE (SilverPlatter on ARC: 1980–November 2001)

Searched: 25 February 2002

1. osteoporosis
2. explode "osteoporosis"/ all subheadings
3. bone near densit*
4. "bone loss*"
5. osteoporotic
6. "bone deminerali*"
7. #1 or #2 or #3 or #4 or #5 or #6
8. "Economics"/ all subheadings
9. explode "Costs-and-Cost-Analysis"/ all subheadings
10. economic value of life in mesh
11. "Economics-Dental"/ all subheadings
12. explode "Economics-Hospital"/ all subheadings
13. "Economics-Medical"/ all subheadings
14. "Economics-Nursing"/ all subheadings
15. "Economics-Pharmaceutical"/ all subheadings
16. (econom$ or cost or costs or costly or costing or price or prices or pricing or pharmacoeconomic$) in ti ab
17. (expenditure$ not energy) in ti ab
18. (value near1 money) in ti ab
19. budget* in ti ab
20. #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19
21. letter in pt
22. editorial in pt
23. #21 or #22
24. #20 not #23
25. animal in tg
26. human in tg
27. #25 not (#25 and #26)
28. (metabolic near cost) in ti ab mesh
29. ((energy or oxygen) near cost) in ti ab mesh
30. #24 not (#27 or #28 or #29)
31. #7 and #30
32. #31 and (LA = "ENGLISH")

EMBASE (SilverPlatter on ARC: 1980–December 2001)

Searched: 25 February 2002

1. "cost-benefit-analysis"/ all subheadings
2. "cost-effectiveness-analysis"/ all subheadings
3. "cost-minimization-analysis"/ all subheadings
4. "cost-utility-analysis"/ all subheadings
5. "economic-evaluation"/ all subheadings
6. cost effect* in ti,ab
7. cost benefit* in ti,ab
8. economic evaluation* in ti,ab
9. technology assessment* in ti,ab
10. pharmacoeconomic* in ti,ab
11. cost util* in ti,ab
12. #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11
13. osteoporosis
14. explode "osteoporosis"/ all subheadings
15. bone near densit*
16. "bone loss*"
17. osteoporotic
18. "bone deminerali*"
19. #13 or #14 or #15 or #16 or #17 or #18
20. #12 and #19
21. LA = "ENGLISH"
22. #20 and (LA = "ENGLISH")

EconLit (SilverPlatter on ARC: 1980–January 2002)

Searched: 25 February 2002

1. osteoporosis
2. explode "osteoporosis"/ all subheadings
3. bone near densit*
4. "bone loss*"
5. osteoporotic
6. "bone deminerali*"
7. #1 or #2 or #3 or #4 or #5 or #6

NHS EED administrative database at CRD (CAIRS)

Searched: 21 February 2002

1. s osteoporosis
2. s bone(2W)densit$
3. s bone(W)loss$
4. s osteoporotic
5. s bone(W)deminerali$
6. s S1 or S2 or S3 or S4 or S5

HTA Database/CRD/CHE CATALOGUE (CAIRS)

Searched: 21 February 2002

1. s osteoporosis
2. s bone(2W)densit$
3. s bone(W)loss$
4. s osteoporotic
5. s bone(W)deminerali$
6. s S1 or S2 or S3 or S4 or S5
7. s Cost$ or econom$ or pharmacoecom$ or pric$ or value or expenditure
8. s S6 and S7

HEED (CD-ROM issue: January 2002)

Searched: 25 February 2002

The following terms were searched in all data fields:

osteoporosis or bone densit* or bone loss* or osteoporotic or bone deminerali*

Health Technology Assessment Programme

Prioritisation Strategy Group

Members

Chair,
Professor Tom Walley,
Director, NHS HTA Programme,
Department of Pharmacology & Therapeutics,
University of Liverpool

Professor Bruce Campbell,
Consultant Vascular & General Surgeon, Royal Devon & Exeter Hospital

Professor Shah Ebrahim,
Professor in Epidemiology of Ageing, University of Bristol

Dr John Reynolds, Clinical Director, Acute General Medicine SDU, Radcliffe Hospital, Oxford

Dr Ron Zimmern, Director,
Public Health Genetics Unit,
Strangeways Research Laboratories, Cambridge

HTA Commissioning Board

Members

Programme Director,
Professor Tom Walley,
Director, NHS HTA Programme,
Department of Pharmacology & Therapeutics,
University of Liverpool

Chair,
Professor Shah Ebrahim,
Professor in Epidemiology of Ageing, Department of Social Medicine, University of Bristol

Deputy Chair,
Professor Jenny Hewison,
Professor of Health Care Psychology, Academic Unit of Psychiatry and Behavioural Sciences, University of Leeds School of Medicine

Dr Jeffrey Aronson
Reader in Clinical Pharmacology, Department of Clinical Pharmacology, Radcliffe Infirmary, Oxford

Professor Ann Bowling,
Professor of Health Services Research, Primary Care and Population Studies,
University College London

Professor Andrew Bradbury,
Professor of Vascular Surgery,
Department of Vascular Surgery,
Birmingham Heartlands Hospital

Professor John Brazier, Director of Health Economics,
Sheffield Health Economics Group, School of Health & Related Research,
University of Sheffield

Dr Andrew Briggs, Public Health Career Scientist, Health Economics Research Centre,
University of Oxford

Professor Nicky Cullum,
Director of Centre for Evidence Based Nursing, Department of Health Sciences, University of York

Dr Andrew Farmer, Senior Lecturer in General Practice,
Department of Primary Health Care, University of Oxford

Professor Fiona J Gilbert,
Professor of Radiology,
Department of Radiology,
University of Aberdeen

Professor Adrian Grant,
Director, Health Services Research Unit, University of Aberdeen

Professor F D Richard Hobbs,
Professor of Primary Care & General Practice, Department of Primary Care & General Practice, University of Birmingham

Professor Peter Jones, Head of Department, University Department of Psychiatry,
University of Cambridge

Professor Sallie Lamb, Research Professor in Physiotherapy/Co-Director, Interdisciplinary Research Centre in Health,
Coventry University

Professor Julian Little,
Professor of Epidemiology,
Department of Medicine and Therapeutics, University of Aberdeen

Professor Stuart Logan,
Director of Health & Social Care Research, The Peninsula Medical School, Universities of Exeter & Plymouth

Professor Tim Peters, Professor of Primary Care Health Services Research, Division of Primary Health Care, University of Bristol

Professor Ian Roberts, Professor of Epidemiology & Public Health, Intervention Research Unit, London School of Hygiene and Tropical Medicine

Professor Peter Sandercock,
Professor of Medical Neurology,
Department of Clinical Neurosciences, University of Edinburgh

Professor Mark Sculpher,
Professor of Health Economics,
Centre for Health Economics,
Institute for Research in the Social Services, University of York

Professor Martin Severs,
Professor in Elderly Health Care, Portsmouth Institute of Medicine

Dr Jonathan Shapiro, Senior Fellow, Health Services Management Centre,
Birmingham

Ms Kate Thomas,
Deputy Director,
Medical Care Research Unit,
University of Sheffield

Professor Simon G Thompson,
Director, MRC Biostatistics Unit, Institute of Public Health,
Cambridge

Ms Sue Ziebland,
Senior Research Fellow,
Cancer Research UK,
University of Oxford

Current and past membership details of all HTA 'committees' are available from the HTA website (www.ncchta.org)

Diagnostic Technologies & Screening Panel

Members

Chair,
Dr Ron Zimmern, Director of the Public Health Genetics Unit, Strangeways Research Laboratories, Cambridge

Ms Norma Armston, Freelance Consumer Advocate, Bolton

Professor Max Bachmann Professor Health Care Interfaces, Department of Health Policy and Practice, University of East Anglia

Professor Rudy Bilous Professor of Clinical Medicine & Consultant Physician, The Academic Centre, South Tees Hospitals NHS Trust

Dr Paul Cockcroft, Consultant Medical Microbiologist/Laboratory Director, Public Health Laboratory, St Mary's Hospital, Portsmouth

Professor Adrian K Dixon, Professor of Radiology, Addenbrooke's Hospital, Cambridge

Dr David Elliman, Consultant in Community Child Health, London

Professor Glyn Elwyn, Primary Medical Care Research Group, Swansea Clinical School, University of Wales Swansea

Dr John Fielding, Consultant Radiologist, Radiology Department, Royal Shrewsbury Hospital

Dr Karen N Foster, Clinical Lecturer, Dept of General Practice & Primary Care, University of Aberdeen

Professor Antony J Franks, Deputy Medical Director, The Leeds Teaching Hospitals NHS Trust

Mr Tam Fry, Honorary Chairman, Child Growth Foundation, London

Dr Edmund Jessop, Medical Adviser, National Specialist Commissioning Advisory Group (NSCAG), Department of Health, London

Dr Jennifer J Kurinczuk, Consultant Clinical Epidemiologist, National Perinatal Epidemiology Unit, Oxford

Dr Susanne M Ludgate, Medical Director, Medical Devices Agency, London

Dr William Rosenberg, Senior Lecturer and Consultant in Medicine, University of Southampton

Dr Susan Schonfield, CPHM Specialised Services Commissioning, Croydon Primary Care Trust

Dr Margaret Somerville, Director of Public Health, Teignbridge Primary Care Trust

Professor Lindsay Wilson Turnbull, Scientific Director, Centre for MR Investigations & YCR Professor of Radiology, University of Hull

Professor Martin J Whittle, Head of Division of Reproductive & Child Health, University of Birmingham

Dr Dennis Wright, Consultant Biochemist & Clinical Director, Pathology & The Kennedy Galton Centre, Northwick Park & St Mark's Hospitals, Harrow

Pharmaceuticals Panel

Members

Chair,
Dr John Reynolds, Clinical Director, Acute General Medicine SDU, Oxford Radcliffe Hospital

Professor Tony Avery, Professor of Primary Health Care, University of Nottingham

Professor Stirling Bryan, Professor of Health Economics, Health Services Management Centre, University of Birmingham

Mr Peter Cardy, Chief Executive, Macmillan Cancer Relief, London

Dr Christopher Cates, GP and Cochrane Editor, Bushey Health Centre

Professor Imti Choonara, Professor in Child Health, University of Nottingham, Derbyshire Children's Hospital

Mr Charles Dobson, Special Projects Adviser, Department of Health

Dr Robin Ferner, Consultant Physician and Director, West Midlands Centre for Adverse Drug Reactions, City Hospital NHS Trust, Birmingham

Dr Karen A Fitzgerald, Pharmaceutical Adviser, Bro Taf Health Authority, Cardiff

Mrs Sharon Hart, Managing Editor, *Drug & Therapeutics Bulletin*, London

Dr Christine Hine, Consultant in Public Health Medicine, Bristol South & West Primary Care Trust

Professor Stan Kaye, Professor of Medical Oncology, Consultant in Medical Oncology/Drug Development, The Royal Marsden Hospital

Ms Barbara Meredith, Project Manager Clinical Guidelines, Patient Involvement Unit, NICE

Dr Frances Rotblat, CPMP Delegate, Medicines Control Agency, London

Professor Jan Scott, Professor of Psychological Treatments, Institute of Psychiatry, University of London

Mrs Katrina Simister, New Products Manager, National Prescribing Centre, Liverpool

Dr Richard Tiner, Medical Director, Association of the British Pharmaceutical Industry

Dr Helen Williams, Consultant Microbiologist, Norfolk & Norwich University Hospital NHS Trust

Therapeutic Procedures Panel

Members

Chair,
Professor Bruce Campbell,
Consultant Vascular and General Surgeon, Royal Devon & Exeter Hospital

Dr Mahmood Adil, Head of Clinical Support & Health Protection, Directorate of Health and Social Care (North), Department of Health, Manchester

Dr Aileen Clarke, Reader in Health Services Research, Public Health & Policy Research Unit, Barts & the London School of Medicine & Dentistry, Institute of Community Health Sciences, Queen Mary, University of London

Mr Matthew William Cooke, Senior Clinical Lecturer and Honorary Consultant, Emergency Department, University of Warwick, Coventry & Warwickshire NHS Trust, Division of Health in the Community, Centre for Primary Health Care Studies, Coventry

Dr Carl E Counsell, Senior Lecturer in Neurology, University of Aberdeen

Dr Keith Dodd, Consultant Paediatrician, Derbyshire Children's Hospital

Professor Gene Feder, Professor of Primary Care R&D, Barts & the London, Queen Mary's School of Medicine and Dentistry, University of London

Professor Paul Gregg, Professor of Orthopaedic Surgical Science, Department of Orthopaedic Surgery, South Tees Hospital NHS Trust

Ms Bec Hanley, Freelance Consumer Advocate, Hurstpierpoint

Ms Maryann L. Hardy, Lecturer, Division of Radiography, University of Bradford

Professor Alan Horwich, Director of Clinical R&D, The Institute of Cancer Research, London

Dr Phillip Leech, Principal Medical Officer for Primary Care, Department of Health, London

Dr Simon de Lusignan, Senior Lecturer, Primary Care Informatics, Department of Community Health Sciences, St George's Hospital Medical School, London

Dr Mike McGovern, Senior Medical Officer, Heart Team, Department of Health, London

Professor James Neilson, Professor of Obstetrics and Gynaecology, Dept of Obstetrics and Gynaecology, University of Liverpool, Liverpool Women's Hospital

Dr John C Pounsford, Consultant Physician, North Bristol NHS Trust

Dr Vimal Sharma, Consultant Psychiatrist & Hon Snr Lecturer, Mental Health Resource Centre, Victoria Central Hospital, Wirrall

Dr L David Smith, Consultant Cardiologist, Royal Devon & Exeter Hospital

Professor Norman Waugh, Professor of Public Health, University of Aberdeen

Expert Advisory Network

Members

Professor Douglas Altman,
Director of CSM & Cancer Research UK Med Stat Gp, Centre for Statistics in Medicine, University of Oxford, Institute of Health Sciences, Headington, Oxford

Professor John Bond,
Director, Centre for Health Services Research, University of Newcastle upon Tyne, School of Population & Health Sciences, Newcastle upon Tyne

Mr Shaun Brogan,
Chief Executive, Ridgeway Primary Care Group, Aylesbury

Mrs Stella Burnside OBE,
Chief Executive, Office of the Chief Executive. Trust Headquarters, Altnagelvin Hospitals Health & Social Services Trust, Altnagelvin Area Hospital, Londonderry

Ms Tracy Bury,
Project Manager, World Confederation for Physical Therapy, London

Mr John A Cairns,
Professor of Health Economics, Health Economics Research Unit, University of Aberdeen

Professor Iain T Cameron,
Professor of Obstetrics and Gynaecology and Head of the School of Medicine, University of Southampton

Dr Christine Clark,
Medical Writer & Consultant Pharmacist, Rossendale

Professor Collette Mary Clifford,
Professor of Nursing & Head of Research, School of Health Sciences, University of Birmingham, Edgbaston, Birmingham

Professor Barry Cookson,
Director, Laboratory of Healthcare Associated Infection, Health Protection Agency, London

Professor Howard Stephen Cuckle,
Professor of Reproductive Epidemiology, Department of Paediatrics, Obstetrics & Gynaecology, University of Leeds

Professor Nicky Cullum,
Director of Centre for Evidence Based Nursing, University of York

Dr Katherine Darton,
Information Unit, MIND – The Mental Health Charity, London

Professor Carol Dezateux,
Professor of Paediatric Epidemiology, London

Mr John Dunning,
Consultant Cardiothoracic Surgeon, Cardiothoracic Surgical Unit, Papworth Hospital NHS Trust, Cambridge

Mr Jonothan Earnshaw,
Consultant Vascular Surgeon, Gloucestershire Royal Hospital, Gloucester

Professor Martin Eccles,
Professor of Clinical Effectiveness, Centre for Health Services Research, University of Newcastle upon Tyne

Professor Pam Enderby,
Professor of Community Rehabilitation, Institute of General Practice and Primary Care, University of Sheffield

Mr Leonard R Fenwick,
Chief Executive, Newcastle upon Tyne Hospitals NHS Trust

Professor David Field,
Professor of Neonatal Medicine, Child Health, The Leicester Royal Infirmary NHS Trust

Mrs Gillian Fletcher,
Antenatal Teacher & Tutor and President, National Childbirth Trust, Henfield

Professor Jayne Franklyn,
Professor of Medicine, Department of Medicine, University of Birmingham, Queen Elizabeth Hospital, Edgbaston, Birmingham

Ms Grace Gibbs,
Deputy Chief Executive, Director for Nursing, Midwifery & Clinical Support Servs, West Middlesex University Hospital, Isleworth

Dr Neville Goodman,
Consultant Anaesthetist, Southmead Hospital, Bristol

Professor Alastair Gray,
Professor of Health Economics, Department of Public Health, University of Oxford

Professor Robert E Hawkins,
CRC Professor and Director of Medical Oncology, Christie CRC Research Centre, Christie Hospital NHS Trust, Manchester

Professor F D Richard Hobbs,
Professor of Primary Care & General Practice, Department of Primary Care & General Practice, University of Birmingham

Professor Allen Hutchinson,
Director of Public Health & Deputy Dean of ScHARR, Department of Public Health, University of Sheffield

Dr Duncan Keeley,
General Practitioner (Dr Burch & Ptnrs), The Health Centre, Thame

Dr Donna Lamping,
Research Degrees Programme Director & Reader in Psychology, Health Services Research Unit, London School of Hygiene and Tropical Medicine, London

Mr George Levvy,
Chief Executive, Motor Neurone Disease Association, Northampton

Professor James Lindesay,
Professor of Psychiatry for the Elderly, University of Leicester, Leicester General Hospital

Professor Rajan Madhok,
Medical Director & Director of Public Health, Directorate of Clinical Strategy & Public Health, North & East Yorkshire & Northern Lincolnshire Health Authority, York

Professor David Mant,
Professor of General Practice, Department of Primary Care, University of Oxford

Professor Alexander Markham,
Director, Molecular Medicine Unit, St James's University Hospital, Leeds

Dr Chris McCall,
General Practitioner, The Hadleigh Practice, Castle Mullen

Professor Alistair McGuire,
Professor of Health Economics, London School of Economics

Dr Peter Moore,
Freelance Science Writer, Ashtead

Dr Andrew Mortimore,
Consultant in Public Health Medicine, Southampton City Primary Care Trust

Dr Sue Moss,
Associate Director, Cancer Screening Evaluation Unit, Institute of Cancer Research, Sutton

Professor Jon Nicholl,
Director of Medical Care Research Unit, School of Health and Related Research, University of Sheffield

Mrs Julietta Patnick,
National Co-ordinator, NHS Cancer Screening Programmes, Sheffield

Professor Robert Peveler,
Professor of Liaison Psychiatry, University Mental Health Group, Royal South Hants Hospital, Southampton

Professor Chris Price,
Visiting Chair – Oxford, Clinical Research, Bayer Diagnostics Europe, Cirencester

Ms Marianne Rigge,
Director, College of Health, London

Dr Eamonn Sheridan,
Consultant in Clinical Genetics, Genetics Department, St James's University Hospital, Leeds

Dr Ken Stein,
Senior Clinical Lecturer in Public Health, Director, Peninsula Technology Assessment Group, University of Exeter

Professor Sarah Stewart-Brown,
Director HSRU/Honorary Consultant in PH Medicine, Department of Public Health, University of Oxford

Professor Ala Szczepura,
Professor of Health Service Research, Centre for Health Services Studies, University of Warwick

Dr Ross Taylor,
Senior Lecturer, Department of General Practice and Primary Care, University of Aberdeen

Mrs Joan Webster,
Consumer member, HTA – Expert Advisory Network

Feedback

The HTA Programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (http://www.ncchta.org) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

We look forward to hearing from you.

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 23 8059 5639 Email: hta@soton.ac.uk
http://www.ncchta.org

ISSN 1366-5278