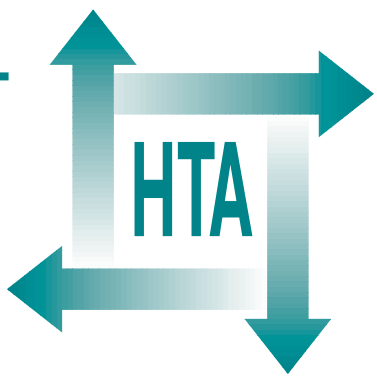
*Health Technology Assessment* 1999; Vol. 3: No. 6

**Review**

# Assessing the costs of healthcare technologies in clinical trials

K Johnston
MJ Buxton
DR Jones
R Fitzpatrick

**Health Technology Assessment**
**NHS R&D HTA Programme**



# Standing Group on Health Technology

## Current members

**Chair:**
**Professor Sir Miles Irving**,
Professor of Surgery, University of Manchester, Hope Hospital, Salford

Professor Martin Buxton,
Professor of Economics, Brunel University

Professor Francis Creed,
School of Psychiatry & Behavioural Sciences, University of Manchester

Professor Charles Florey,
Department of Epidemiology & Public Health, Ninewells Hospital & Medical School, University of Dundee

Professor John Gabbay,
Director, Wessex Institute for Health Research & Development

Professor Sir John Grimley Evans,
Department of Geriatric Medicine, Radcliffe Infirmary, Oxford

Dr Tony Hope,
The Medical School, University of Oxford

Professor Richard Lilford,
Regional Director, R&D, West Midlands

Dr Jeremy Metters,
Deputy Chief Medical Officer, Department of Health

Professor Maggie Pearson,
Regional Director of R&D, NHS Executive North West

Mr Hugh Ross,
Chief Executive, The United Bristol Healthcare NHS Trust

Professor Trevor Sheldon,
Director, NHS Centre for Reviews & Dissemination, University of York

Professor Mike Smith,
Director, The Research School of Medicine, University of Leeds

Dr John Tripp,
Department of Child Health, Royal Devon & Exeter Healthcare NHS Trust

Professor Tom Walley,
Department of Pharmacological Therapeutics, University of Liverpool

Dr Julie Woodin,
Chief Executive, Nottingham Health Authority

Professor Kent Woods
(**Chair Designate**),
Regional Director of R&D, NHS Executive, Trent

## Past members

Dr Sheila Adam,
Department of Health

Professor Angela Coulter,
Director, King's Fund, London

Professor Anthony Culyer,
Deputy Vice-Chancellor, University of York

Dr Peter Doyle,
Executive Director, Zeneca Ltd, ACOST Committee on Medical Research & Health

Professor John Farndon,
Professor of Surgery, University of Bristol

Professor Howard Glennester,
Professor of Social Science & Administration, London School of Economics & Political Science

Mr John H James,
Chief Executive, Kensington, Chelsea & Westminster Health Authority

Professor Michael Maisey,
Professor of Radiological Sciences, Guy's, King's & St Thomas's School of Medicine & Dentistry, London

Mrs Gloria Oates,
Chief Executive, Oldham NHS Trust

Dr George Poste,
Chief Science & Technology Officer, SmithKline Beecham

Professor Michael Rawlins,
Wolfson Unit of Clinical Pharmacology, University of Newcastle-upon-Tyne

Professor Martin Roland,
Professor of General Practice, University of Manchester

Professor Ian Russell,
Department of Health Sciences & Clinical Evaluation, University of York

Dr Charles Swan,
Consultant Gastroenterologist, North Staffordshire Royal Infirmary

Details of the membership of the HTA panels, the NCCHTA Advisory Group and the HTA Commissioning Board are given at the end of this report.



**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.

# Assessing the costs of healthcare technologies in clinical trials

K Johnston[1]
MJ Buxton[1]
DR Jones[2]
R Fitzpatrick[3]

[1] Health Economics Research Group, Brunel University, Uxbridge, UK
[2] Department of Epidemiology and Public Health, University of Leicester, UK
[3] Division of Public Health and Primary Health Care, University of Oxford, UK

Published April 1999

This report should be referenced as follows:

Johnston K, Buxton MJ, Jones DR, Fitzpatrick R. Assessing the costs of healthcare technologies in clinical trials. *Health Technol Assess* 1999;**3**(6).

*Health Technology Assessment* is indexed in *Index Medicus*/MEDLINE and *Excerpta Medica*/EMBASE. Copies of the Executive Summaries are available from the NCCHTA web site (see overleaf).

# NHS R&D HTA Programme

The overall aim of the NHS R&D Health Technology Assessment (HTA) programme 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 work in the NHS. Research is undertaken in those areas where the evidence will lead to the greatest benefits to patients, either through improved patient outcomes or the most efficient use of NHS resources.

The Standing Group on Health Technology advises on national priorities for health technology assessment. Six advisory panels assist the Standing Group in identifying and prioritising projects. These priorities are then considered by the HTA Commissioning Board supported by the National Coordinating Centre for HTA (NCCHTA).

This report is one of a series covering acute care, diagnostics and imaging, methodology, pharmaceuticals, population screening, and primary and community care. It was identified as a priority by the Methodology Panel and funded as project number 93/48/05.

The views expressed in this publication are those of the authors and not necessarily those of the Standing Group, the Commissioning Board, the Panel members or the Department of Health. The editors wish to emphasise that funding and publication of this research by the NHS should not be taken as implicit support for the recommendations for policy contained herein. In particular, policy options in the area of screening will be considered by the National Screening Committee. This Committee, chaired by the Chief Medical Officer, will take into account the views expressed here, further available evidence and other relevant considerations.

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 editors have tried to ensure the accuracy of this report but cannot accept responsibility for any errors or omissions. They would like to thank the referees for their constructive comments on the draft document.

ISSN 1366-5278



Published by Core Research, Alton, on behalf of the NCCHTA.
Printed on acid-free paper in the UK by The Basingstoke Press, Basingstoke.

Copies of this report can be obtained from:

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 1703 595 639 Email: hta@soton.ac.uk
http://www.soton.ac.uk/~hta

# Contents

# Executive summary

## Background

In the economic evaluation of healthcare technologies, costs are estimated by multiplying the quantities of resources used by the unit costs of the resources. When economic evaluations are conducted alongside clinical trials, the opportunity arises to collect comprehensive and detailed information on resource-use quantities. For example, resource use such as days in hospital can be measured for each individual in the trial. This then allows the estimation of cost data at the individual level, referred to as 'patient-specific' data. The advantage of such data is that it allows statistical analysis of costs to be performed.

There is, however, a legitimate concern not to overburden the trial data collection process with the gathering of such detailed resource-use information. Consequently, the choice of resource-use items for data collection needs to be considered very carefully. This report identifies and examines the range of methodological issues concerning the collection of resource-use data for costing purposes and its analysis.

## Objectives

The overarching objective is to challenge investigators to think through their study design in order to collect appropriate resource-use information in the most efficient way. Specifically, the objectives are:

- to identify methodological issues concerning the collection of resource-use data for costing purposes and its analysis
- to classify methodological issues into: (1) those where there is general agreement about how they should be handled; (2) those remaining open because of legitimate differences in values or perspectives; and (3) those where further empirical testing could resolve how the issue should be handled
- to demonstrate how existing data can be used to inform the design of costing studies in trials
- to develop a framework or decision aid within which decisions about costing in specific trials can be made.

## Methods

The methodological issues were identified through a review of several strands of relevant literature, including methodological review articles, empirical articles and guidelines on performing economic evaluations. In developing the review, comments from relevant experts were sought with the aim of identifying further issues and opinions. The methodological issues identified are structured under four broad headings:

- study design
- data collection
- data analysis
- presentation of results.

The two final objectives listed above are achieved through empirical analysis and the development of a framework or decision aid, respectively. Further detail on the methods is presented in the main report.

## Results

Design issues address the types of cost to be included, such as health service, trial, future and productivity costs. The decision on which types of cost to include depends on seven key factors:

- possible links to economic welfare theory
- the perspective to be adopted
- the form of economic evaluation
- the avoidance of double counting
- the quantitative importance of the type of cost
- whether the cost can be attributed to the intervention
- the time horizon of the study.

The collection of detailed data on resource use for all patients may not be necessary; key cost-generating events can be measured. These can be defined as where there is variation in the frequency of events between arms of the trial or between patients within arms. Determining sample sizes for detecting differences in costs or cost-effectiveness involves identifying an economically important difference and having information on the variability of cost data from previous studies or from

pilot studies. A further sampling issue to be addressed in multicentre trials is the selection of centres and whether resource-use and unit cost information should be collected from all centres.

Data collection issues involve deciding on the appropriate resource-use data collection method. Resource-use data can be measured on a patient-specific basis by using, for example, interviews, questionnaires, case record forms or diary cards. In selecting a method, potential sources of bias have to be addressed, including recall bias, evasive answer bias, non-response bias, selection bias and question format. The validity and reliability of resource-use data collection methods have not been tested fully and are therefore not reported in the literature.

Data analysis may also influence the design of the study. In summarising and synthesising cost data, issues such as how to pool data and how to handle missing and censored data have to be addressed. It is important to take into account the variability in cost data and its distribution. It is generally agreed that mean costs convey more useful information than medians because they relate to total cost. The methods used to address uncertainty in methods and results include both statistical and sensitivity analyses; these have complementary roles. Sensitivity analysis can also be used to generalise results.

The presentation of results addresses reporting formats. Results should be presented in a disaggregated manner, for example, by separating resource use from unit costs and reporting the contribution of different types of cost to total costs. The development of a common reporting format for economic evaluations would increase the transparency of both methods and results. The design of future studies relies on transparent reporting in earlier studies so that issues such as the variability in cost data can be determined.

There are two additional elements of the review. First, an existing data set on costing from a clinical trial was used to illustrate how evidence relating to costs from a completed study can be used to inform the design of data for costing. By examining the results of detailed data collection, the exercise illustrates that, in the example at least, it is possible for simpler data collection methods to be adopted to produce comparable results. The exercise demonstrates the usefulness of having access to, and analysing, existing data sets in order to address design issues.

Secondly, a decision aid, or structured framework, has been developed within which decisions can be made about designing a costing study alongside a clinical trial. In effect, the decision aid requires answering a set of explicit questions. It is recommended that it should be tested in future studies.

## Conclusions

Methodological issues on which there is general agreement include identifying perspective, measuring units of resource use, and applying appropriate unit cost. Those issues remaining open because of legitimate differences in values or perspectives concern which perspective to adopt and whether to base decisions on economic welfare theory. Finally, methodological issues requiring further empirical study include:

- exploring optimal sampling approaches
- questions surrounding multicentre clinical trials
- testing the validity and reliability of resource-use data collection methods
- handling missing and censored data
- methods used to generalise results.

By presenting issues in this way, the review recognises the inevitability of some issues remaining unresolved while at the same time allowing the specification of a future research agenda.

## Recommendations

Four sets of recommendations are provided, for: investigators, funding bodies, those responsible for ensuring high standards in reporting of studies, and further research. The review and its associated appendices serve to challenge investigators to think through methodological issues and to decide how best they can be handled in their own circumstances. For those issues requiring further empirical investigation, researchers should build empirical testing into their studies. In this way, methodological standards in the next generation of studies can be improved, and the future research necessary to develop further and refine methodology can be undertaken. In the short term, however, the review will provide users of currently available studies with information having a critical basis against which to assess the cost information presented.

# Chapter 1
# Introduction

## Background

The relatively recent growth of serious interest in the formal economic evaluation of health technologies has led to the recognition that higher and more consistent methodological standards are required to ensure that the results available to decision makers are appropriate, reliable and comparable. In addition, there is concern that economic evidence should itself be obtained in a cost-effective manner.

In particular, because economic analyses are being proposed more frequently, or are required alongside or in relation to clinical trials, a range of issues arise, of both methodological principle and best practice, for the process of obtaining and analysing data (Drummond and Davies, 1991). In this context, there is an additional set of considerations concerning how cost data collection can best relate to the gathering of data on clinical and quality of life end-points (Dubois *et al.,* 1993).

Many of the outstanding issues of methodology cannot in the end be resolved simply by reference to scientific evidence or methodological principles, but reflect normative questions about what information different users of economic evidence need in particular circumstances. Certainly, recourse to determining methodological choices by review of past practice, however systematic that review, has the danger of simply perpetuating past misjudgements. A further danger is that premature definition of best practice may effectively lead to a freezing of the methodological development that would otherwise continue.

There are already a number of more or less authoritative statements of methodological principle (Drummond *et al.,* 1987, 1997; Detsky and Naglie, 1990; Luce and Elixhauser, 1990a,b; Gold *et al.,* 1996; Russell *et al.,* 1996; Weinstein *et al.,* 1996; Donaldson and Shackley, 1997). There is also literature on issues arising when performing economic evaluations alongside clinical trials (Drummond and Stoddart, 1984; Bulpitt *et al.*, 1990a,b; Drummond and Davies, 1991; Adams *et al.*, 1992; Bennett *et al.*, 1994a,b; Drummond, 1994, 1995). Other studies have focused exclusively on costing methods in general (Simpson and Souney, 1988; Donaldson, 1990; Evans, 1990; Johannesson, 1994; Jacobs and Bachynsky, 1996; Wolff *et al.*, 1997). In addition, a number of context-specific sets of guidance exist (Association of the British Pharmaceutical Industry, 1994; Canadian Coordinating Office for Health Technology Assessment, 1994; Ministry of Health, 1994; Belgian Society for Pharmacoepidemiology, 1995; Commonwealth of Australia, 1995). Although all these works are important as statements of agreed principle and to highlight points of current disagreement, they do not readily provide practical advice to those who are designing, or reviewing the design of, clinical trials that will provide economic information relevant to the NHS. This review aims to fill that perceived gap of practical advice.

## Definitions

**Cost** is the product of two elements: the quantity of resources consumed and the unit cost of the resources. This type of cost is sometimes referred to as **resource cost**. Resource use, for example, a day in hospital, is also referred to as a **cost-generating event**. Cost-generating events may be measured on a **patient-specific** or **non-patient-specific** basis. Events, or costs, that are patient-specific are **stochastic**; that is, they vary in number and frequency from patient to patient. Events, or costs, that are non-patient-specific are **deterministic**; that is, they are the same for each patient. **Unit costs**, such as staff costs per hour or the cost of an inpatient stay, are sometimes referred to in the literature as values of resource-use or unit prices, but the term 'unit cost' is used here. **Opportunity cost** is the theoretical concept of cost used in economic evaluation and refers to the benefit that could have been obtained from the next best use of resources. The **average cost** of a unit of service is the total cost divided by the total number of units. The **marginal cost** is the cost of providing an extra unit of a service. A **cost–consequence analysis** presents the costs and consequences separately in a disaggregated form. **Cost-effectiveness analysis** presents its results as an incremental ratio of additional costs to additional effects for one treatment or intervention compared with another, for example, the additional cost per life year gained when comparing one intervention with another.

The **numerator** of the ratio consists of costs and the **denominator** of effects. **Cost–utility analysis** produces an incremental ratio of additional costs to additional quality adjusted life years gained. **Cost–benefit analysis** presents results as a net monetary benefit or cost.

## Overview of alternative approaches to costing

The focus of this review is costing within the context of clinical trials. The key advantage of using clinical trials as a framework for economic evaluation is that they provide the opportunity to collect and analyse patient-specific resource-use (and hence cost) data. Clearly, clinical trials are not the only study design able to generate patient-specific resource-use data, but this focus was chosen for two reasons: first, because there is a growing number of published (and commissioned) economic evaluations conducted as part of clinical trials; and secondly, because costing studies based on other study designs raise issues that are a subset of those raised by clinical trials. Approaches to costing and their relationships to study designs are now discussed.

Alternative approaches to costing can be identified according to whether or not resource use is measured on a patient-specific basis. If a cost analysis is based on patient-specific data then it is stochastic because resource use varies by patient. This means that statistical analysis of resource-use and cost data can be performed. If a costing study is based on resource-use data that is non-patient-specific, then it is deterministic; that is, resource use is assumed to be the same for patients receiving the same intervention.

Even where resource use can be measured on a patient-specific basis, it may not be necessary to measure all resource use on this basis for two reasons. First, if there is no variation in some element of resource use between patients, then a decision may be taken to measure only those resources that are likely to vary between patients on a patient-specific basis. This decision may be based on a number of factors that are discussed in this review. Secondly, some resource use cannot be identified easily at a patient-specific level; for example, the use of buildings and overheads, such as heating and electricity. This review addresses issues in measuring both patient-specific and non-patient-specific resource-use data.

Building on this distinction between patient- and non-patient-specific resource-use data, costing studies and, thus, economic evaluations, can be classified into three broad approaches (O'Brien *et al.*, 1994; Drummond *et al.*, 1997), as summarised in *Table 1*.

Even within the third approach, only some of the resource use may be measured on a patient-specific basis for the reasons outlined above. This classification ignores the role of unit cost information and the fact that, in most cases, deterministic unit costs are attached to either patient-specific (stochastic) or non-patient-specific (deterministic) resource use.

Within these alternative costing approaches, different study designs may be used to generate effectiveness data. These include:

- experimental designs such as clinical trials
- observational studies such as cohort studies and multiple case series
- studies involving the synthesis of existing data or results from multiple sources, such as meta-analyses or systematic reviews.

In the latter type of study design, effectiveness data are combined and sometimes synthesised using a decision analytical or modelling framework. The nature of the effectiveness data will

***TABLE 1*** *Approaches to costing*[a]

| Costing approach | Measurement of resource-use and effectiveness data |
|---|---|
| Deterministic | Non-patient-specific resource-use data<br>Non-patient-specific effectiveness data |
| Partially stochastic | Non-patient-specific resource-use data<br>Patient-specific effectiveness data |
| Wholly stochastic | Patient-specific resource-use data<br>Patient-specific effectiveness data |

[a] *Adapted from O'Brien* et al., *1994; Drummond* et al., *1997*

have implications for the type of resource-use data available. For example, a design requiring synthesis of existing data will generate non-patient-specific resource-use data. Some study designs raise additional issues for resource-use data collection.

- Clinical trials may limit the generalisability of results but this is an issue that is equally relevant for other study designs and is included in the review.
- Cohort studies may not make comparisons between groups; thus, sampling issues are different for this type of study design.

In general, the methodological issues in costing are common across study designs and the review is therefore inclusive in considering relevant methodological issues for all types of study design.

## Purpose and scope of the systematic review

This systematic review focuses on issues concerning the collection and analysis of data, for the purposes of costing within the context of clinical trials. The main objective is to review available conceptual and empirical evidence in order to establish how more informed and appropriate choices could be made in the future about study design, data-collection procedures and the analysis of costing data alongside clinical trials.

The review considers the various questions that commonly arise in designing costing studies (summarised in *Box 1*).

**BOX 1 Questions commonly arising when designing costing studies**

- which costs to consider
- in what form to collect the data
- how to determine sample size and the sampling approach
- how and when to collect the data
- how to analyse cost data and relate it to data on effectiveness and quality of life.

The review's target audience includes those who are designing clinical trials and those reviewing designs, either prospectively as part of the research commissioning process or retrospectively in judging evidence. The target audience is multi-disciplinary and thus the language used and the presentation aims to reflect this. The review will benefit the NHS in the short term in providing users of currently available information with a critical basis against which to assess the cost information presented; in the medium term by utilising current knowledge fully, to improve the methodological standards to which the next generation of studies are conducted; and, in the longer term, by identifying the most important further research necessary to develop and refine methodology.

Although the focus of the review is on costing alongside clinical trials, the methodological issues were identified through a systematic review of several strands of relevant literature (summarised in *Box 2*).

**BOX 2 Strands of literature upon which the systematic review is based**

- literature reviews of the specific topic
- empirical articles conducting economic evaluations alongside clinical trials
- literature reviews on costing methods in economic evaluation in general
- guidelines on performing economic evaluations.

First, reviews of the specific topic (costing alongside clinical trials) were included, as well as review articles addressing single methodological issues. Secondly, empirical papers conducting economic evaluations alongside clinical trials were included because they might raise additional methodological issues. The relatively recent increase in economic evaluations conducted alongside clinical trials meant that many of the methodological issues might be raised in recent empirical studies. Thirdly, the literature on costing methods in economic evaluation in general was included because this forms the basis for the design of many economic evaluations. Fourthly, guidelines recently produced for various countries, which attempt to standardise methodologies employed for economic evaluation, were included because they make recommendations about methods. Finally, guidelines for authors intending to publish economic evaluations in peer reviewed journals were also included in the review since, by implication, they comment on methodological issues. The search strategy for the systematic review is presented in detail in appendix 1.

The literature was reviewed with the aim of identifying methodological principles and issues relating to the collection and analysis of cost data within the context of clinical trials. The aim was not to record the frequencies with which issues were

raised but to be comprehensive in identifying methodological issues; nor was the aim to provide a set of guidelines. Instead, the aim was to distinguish different types of methodological issue:

- those where there is general agreement
- those where there is disagreement and which remain open because of legitimate differences in values and perspectives
- those where there is disagreement that could be resolved with further research and empirical testing.

In developing the review, structured comments from relevant experts were sought with the aim of identifying further issues and opinions. This process is described in appendix 1.

The review does not comment in detail on the derivation of values for resources (i.e. unit costs) because methods for their estimation relate to different publications. Issues surrounding the unit costs of resources are not, however, totally excluded since the availability of unit costs may constrain the way in which resource use can be appropriately recorded. Options for sources of unit cost data are, therefore, discussed. For similar reasons, the review does not draw on the literature surrounding the rate at which costs are discounted, but points the reader to further literature on this topic. The ultimate goal is to challenge investigators to think through their hypotheses with regard to cost and how to collect cost information in the most efficient manner in order to achieve their study goals. The aims of the review are summarised in *Box 3*. The work relating to the final two aims is reported in appendices 2 and 3. Clinical trials afford the opportunity to collect comprehensive and detailed data but in practice there is a legitimate concern not to overburden the trial data collection process. Consequently, the choice of resource-use items and methods needs to be carefully considered.

**BOX 3 Summary of aims of systematic review**

- to identify relevant methodological issues
- to classify methodological issues into: (1) those where there is general agreement about how they should be handled; (2) those remaining open because of legitimate differences in values or perspectives; and (3) those where further empirical testing could resolve how the issue should be handled
- to challenge investigators to think through their study design in order to collect cost information in the most efficient way
- to demonstrate how existing data can be used to inform the design of costing studies in trials
- to develop a framework within which decisions about costing in specific trials can be made.

## Structure of the systematic review

The review is divided into six sections. Following this introduction, chapter 2 discusses issues of study design in general. Those relating to the design of data collection methods are described in chapter 3. This is followed by discussion of the issues surrounding the analysis and presentation of results in chapters 4 and 5. The final chapter discusses the findings of the review and makes recommendations for investigators and future research. Where issues in different chapters relate to each other, this is cross-referenced in the text.

In appendix 1, the methods used to identify and select articles for the systematic review are described in detail. Appendix 2 presents an empirical demonstration of how existing data sets can be used to inform the design of costing studies. Building on the systematic review of methodological issues, appendix 3 proffers a decision aid to assist in designing cost data collection in clinical trials.

# Chapter 2

# Study design

## Types of cost

Traditionally, the types of cost for potential inclusion in an economic evaluation have been classified into five main groups (Drummond *et al.*, 1987; Luce and Elixhauser, 1990a; Weinstein, 1990):

- direct healthcare costs (for example, hospital care, drug use)
- direct non-healthcare costs (for example, patient travel costs)
- indirect healthcare costs (the costs of healthcare consumption during years of life gained as a result of a healthcare intervention)
- indirect non-healthcare costs (the value of production loss due to illness or treatment; the opportunity cost of time spent)
- intangible costs (the pain and suffering associated with treatment).

More recently, there has been a move away from classifying costs into 'direct' and 'indirect' to a more detailed categorisation (Gold *et al.*, 1996; Drummond *et al.*, 1997). These are discussed below. Indirect healthcare costs have also been replaced by a more detailed categorisation of what are now more commonly referred to as 'future costs'. For example, recent literature has classified future costs into several categories according to both whether they are related to the intervention in question and when they occur (Gold *et al.*, 1996; Meltzer, 1997). The term 'intangible' is less frequently used. Costs formerly classified under this heading had the potential to be double counted, that is, measured both as a cost and as a (dis)benefit.

For the purposes of this review, because no common agreement was found in the literature regarding classification of costs, they are classified into three broad categories: health service costs, non-health service costs and non-resource costs. *Box 4* summarises this classification of cost and the types of cost falling into each category. These costs are defined and described in the remainder of this section. Within each type, decisions have to be made about whether their inclusion is appropriate. A general discussion of the basis upon which the choice of costs for inclusion can be made is presented in the next section of this chapter (pp. 9–12). Where specific issues of inclusion or exclusion arise for an individual type of cost, they are raised when discussing that type of cost. Methodological issues that arise relating to both the measurement and valuation of the different types of cost are discussed later in this chapter (pp. 14, 18).

**BOX 4 Summary of types of cost**

**Health service costs**
- direct costs of the whole intervention (e.g. hospital care)
- general illness costs (e.g. costs of treating illnesses other than the one associated with the intervention)
- future health service costs
- trial costs.

**Non-health service costs**
- costs incurred by other public sector budgets (e.g. social services)
- patient's travel costs
- other out of pocket expenses incurred by the patient
- informal care costs
- patient's time costs incurred while receiving treatment
- productivity costs associated with morbidity and mortality
- future non-health service costs.

**Non-resource costs**
- transfer payments.

### Health service costs

**Direct costs of the whole intervention** include staff time, drug use and other medical supplies. They also include hospital resources, such as the number of treatments and bed days, as well as outpatient, general practitioner (GP) and nurse visits. Finally, they include the use of buildings, other capital and equipment, and overheads, such as heating and lighting, arising from the health service intervention.

The term 'whole intervention' is used to stress the fact that the costs of the intervention should include the broader health service costs. For example, if a screening programme is being evaluated, they should include the cost of the

screening visits as well as other related health service visits, such as GP visits. In a study of cardiovascular screening, Wonderling and co-workers (1996b) included the costs of other health service visits in addition to the costs of screening visits, thereby costing the health service resources associated with the whole intervention. Issues of attribution arise, for example, of whether attendance at a GP's surgery for depression is related to the initial treatment. Attribution of costs is discussed further in the next section of this chapter (p. 11).

**General illness costs** are the costs of undergoing therapy for other illnesses while being treated for the intervention in question. These may be illnesses arising from the intervention or they may be unrelated to it. These are costs that occur during the trial period rather than in the future. Whether a cost is attributable to an intervention is discussed below (p. 11).

**Future health service costs** are the additional costs of treatment for diseases arising as a result of individuals living longer because of the initial intervention (Gold *et al.*, 1996; Meltzer, 1997). These costs can be further classified by whether they arise from **related** or **unrelated** diseases or treatments, as well as by whether they occur in **years of life lived anyway** or in **years of life gained** because of the original treatment (Gold *et al.*, 1996) (*Box 5*). Many cost-effectiveness studies include future costs for related illnesses only, ignoring costs for unrelated illnesses (Meltzer, 1997). Defining precisely what is meant by 'related' is difficult; it has been taken to mean both disease-specific costs and costs of other diseases (Meltzer, 1997). Again, whether or not a cost is related is a question of attribution, as discussed in the next section (p. 11).

**BOX 5 Classification of future health service costs**[a]

- related, occurring in years of life lived anyway
- unrelated, occurring in years of life lived anyway
- related, occurring in years of life gained
- unrelated, occurring in years of life gained.

[a] *Adapted from Gold* et al., *1996; Meltzer, 1997*

There is no consensus in the literature about whether all types of future costs should be included. Gold and co-workers (1996) argue that future costs for related diseases incurred in **years of life that would have been lived anyway** should be included in economic evaluation. These are costs that occur in years of life that would have been lived without the intervention, for example, the costs of treating heart attacks, if these are affected by the intervention.

Gold and colleagues (1996) argue further that future costs associated with costs for unrelated diseases incurred in years of life that would have been lived anyway should be omitted because they are the same with or without the intervention and therefore cancel out in the calculation of the incremental cost-effectiveness ratio. They also argue that including them may induce errors in the estimation of costs because of variability in unrelated costs with and without the intervention. This argument could, however, be applied to any form of cost.

Related future costs in **years of life gained** (i.e. an individual living longer because of an intervention) occur when, for example, a coronary artery bypass graft delays a myocardial infarction by 5 years. Gold and co-workers (1996) argue that the costs of treating all coronary events should be included in the 5-year period. Future costs in years of life gained occur commonly in prevention programmes, where treatment is often delayed. It is usual to include all the future effects of the prevention programme in the economic evaluation and therefore, for consistency, it could be argued that future costs should be included (Drummond *et al.*, 1997). These types of costs are therefore often included.

The element of future costs where there is less agreement is whether to include health service costs for unrelated diseases that occur in years of life gained. Mushlin and Fintor (1992) argue that any additional costs for the treatment of unrelated disease arising because individuals live longer owing to screening and early treatment of breast cancer should be excluded, but the basis of this argument is not clear. Johannesson and colleagues (1997) argue that the difference between consumption and production during life years gained should be included. Morris and co-workers (1997) argue, in a review of the cost-effectiveness of strategies for preventing hypercholesterolaemia, that inclusion of the higher costs of routine care for non-cardiovascular disease accruing in years of life gained are only relevant if total mortality rather than cardiac heart disease mortality is the basis of the effectiveness measure. Meltzer (1997) argues that excluding future costs biases cost-effectiveness estimates in favour of interventions that increase length of life rather than quality of life and therefore they should be included. This raises the important point that unrelated future costs in life years gained are not independent of age and, if interventions add years of life for different age

groups, this will affect the cost-effectiveness ratio (Gold *et al.*, 1996). Unrelated future costs in life years gained may be small compared with other costs in the analysis but should only be excluded if the preferred approach is to base the decision on the types of cost to include on quantitative importance (discussed further later in this chapter; p. 11).

The basis on which the arguments for including or excluding future health service costs can be summarised is as follows. First, the difference between consumption and production during life years gained is an argument for inclusion. Secondly, consistency of cost measurement with effectiveness measurement is important. This implies that only if total mortality is the measure of effectiveness would future unrelated health service costs in life years gained be included. Restricting effectiveness to disease-specific measures ignores the possibility of interventions for one condition affecting another, and that some costs may be related to an intervention but may not be disease specific. This relates to the attribution of costs discussed in the next section (p. 11). Thirdly, quantitative importance may be a relevant criterion; if there is likely to be no difference between interventions in terms of unrelated future health service costs in life years gained then they could be omitted. For each classification of future cost, however, there is a subjective element in the decision about whether to include it, since it is not known in advance whether a cost will be related or unrelated. Given the disagreement concerning how to handle future health service costs, analysts should consider the impact on the results of the inclusion or exclusion of future costs through methods such as sensitivity analysis (discussed in chapter 4; p. 31–32).

**Trial costs** are a further type of cost. They include the costs of carrying out the research, rather than the cost of treatment. They also refer to the costs of procedures in the trial that would definitely not be performed in routine practice and are therefore required solely for the purposes of the trial itself. These events may arise because patients are more closely monitored in the trial or because of the necessity to preserve blinding (Drummond and Davies, 1991). Related to this are protocol-driven costs, which usually refer to events for which the timing and/or frequency is set by the trial protocol and hence does not vary between the arms of the trial but which, in a real world situation, may be required to vary.

Including trial costs in the cost analysis may over- (or under-)estimate the true costs (Rittenhouse, 1996b). In general, it is recommended that these costs are excluded (Canadian Coordinating Office for Health Technology Assessment, 1994). Langham and colleagues (1996) removed costs that were determined by the trial protocol. In their study these were the costs of re-examination in a cardiovascular screening programme. Similarly, in a study comparing treatments for asthma, Rutten-van Molken and co-workers (1995) excluded the costs of special visits that were required for the study protocol only. Wonderling and colleagues (1996b), in a study of cardiovascular screening, included only those trial resources that would have been required to provide the service routinely.

There are several important issues to consider in deciding whether to include or exclude trial costs. First, it is not always straightforward to separate service costs from trial costs (Wonderling *et al.*, 1996a). Secondly, trial costs may impact on outcome and on this basis it has been argued that they should be included (Drummond and Jefferson, 1996). Trial costs can only be ignored if they are known not to affect outcome. Thirdly, how these costs should be handled will depend on the extent to which treatment in the trial reflects routine practice. The more pragmatic the trial design, the fewer the costs that are protocol driven, the more resource-use estimates are likely to reflect actual practice. Analysts should therefore seek to minimise protocol-driven costs. This issue also relates to the generalisability of results from trials, which is discussed later in this chapter (p. 19) and in chapter 4 (p. 32).

#### *Applied example of health service costs*

An example of health service costs within a study context is presented in *Box 6*. It may be difficult to distinguish the costs arising from illness from the costs of the intervention.

## Non-health service costs

Non-health service costs include social services costs, and patients' travel and time costs. In deciding whether or not to include non-health service costs, an important consideration is whether the intervention under study shifts the cost burden to other public sector budgets or to the patient. If cost shifting is possible, then it is important to take such costs into account.

**Costs incurred by other public sector budgets**, such as social services costs, are particularly important for evaluating programmes of care for older people or those with a mental illness. In practice, the boundaries between which costs fall on the health service budget and which fall on social services budgets is often unclear and may change as budgetary arrangements change. If it is unclear which budget the costs

**BOX 6 Example of types of health service costs**

**Study context**

A randomised cardiovascular screening and intervention programme led by practice nurses aimed to achieve reductions in blood pressure, cholesterol concentration and smoking prevalence among participants in the intervention arm and thus reduce subsequent heart disease and stroke. Examples of the different types of health service cost in this context are:

- direct costs of the whole intervention
  - programme costs (nurse time, consumables, buildings costs)
  - drug costs
  - broader health service costs (GP health checks, other health checks)
  - hospitalisations due to heart disease.
- general illness costs
  - costs of treating other illnesses arising from the intervention (e.g. where a visit to a practice nurse identified other illnesses for which treatment was required)
  - costs of treating other illnesses unrelated to the intervention (e.g. inpatient costs for an unrelated accident).
- future costs
  - related costs arising in years of life lived anyway (e.g. the costs associated with the treatment of stroke, if these are affected by the intervention)
  - related costs arising in life years gained (e.g. the costs of treating all coronary events in the life years gained)
- trial costs
  - costs of the research team and the costs of any tests undertaken only for the purposes of the research.

fall under, the implication is that cost shifting may be possible (i.e. costs incurred by one sector may be shifted to another). This is a strong argument for including non-health service costs in a study.

Non-health service costs also include **patients' travel costs** to hospital for treatment (e.g. bus, train and taxi fares and car costs). For example, in a study of radiotherapy treatments for head and neck cancer, the new treatment, continuous hyperfractionated accelerated radiotherapy (CHART), was known to require patients to travel less frequently than the conventional treatment and thus travel costs were an important component (Coyle and Drummond, 1997). **Other out of pocket expenses** may also be incurred by patients as a result of treatment, such as child care costs.

**Informal care costs** are those incurred by family members or friends in caring for patients, usually unpaid. These costs include the financial outlay incurred by the care but also the time spent by the informal carer in providing care. Smith and Wright (1994) argue that studies attempting to include these costs have used inconsistent methodologies. Again, it is important to include this type of cost if an intervention has the potential to shift the burden of costs to informal carers.

The opportunity **cost of patients' time** associated with receiving treatment is a further type of non-health service cost. These time costs have often, confusingly, been referred to as indirect costs (Luce and Elixhauser, 1990a; Weinstein, 1990). They may reflect two different costs, depending on the perspective (see the next section of this chapter; p. 10). First, if a patient or individual perspective is adopted, this may reflect the loss of time to the individual in attending for treatment, whether this is work time or non-work time. These time costs are important if they affect the demand for health services. For example, they may be particularly important in screening programmes where, if time costs are perceived to be too high, they may be a barrier to attendance and ultimately affect the efficiency of the programme (Sculpher and Buxton, 1993; Torgerson *et al.*, 1994). Secondly, if a societal perspective is adopted, time costs incurred by individuals in receiving treatment reflect the loss of production to society. Unlike the productivity costs associated with morbidity discussed below, there is agreement that patients' time costs that are associated with treatment should be included in the numerator (costs) of a cost-effectiveness ratio.

**Productivity costs** may be separated into three phases: treatment (as just discussed); morbidity (incurred as a result of patients being ill); and mortality (incurred as a result of death) (Gold *et al.*, 1996). As with patients' time costs, the production costs associated with patient **morbidity** as a result of treatment can be separated into: effects on the patient (time off work, loss of leisure time); effects on the employer (training and replacement costs); and effects on society (production losses, whether paid or unpaid) (van Roijen *et al.*, 1996; Brouwer *et al.*, 1997a,b). Even if patients take time off from work while recovering from treatment, actual production may not be affected because of the replacement of workers and short-term absences (Koopmanschap and Rutten 1993, 1996; Koopmanschap *et al.*, 1995). Productivity may also be compensated for if losses are avoided by returning to work (the welfare effect of treatment) (Drummond *et al.*, 1997). Production costs associated with **mortality** affect families of the deceased, the employer (training and replacement costs) and society (production losses, whether paid or unpaid).

Although conceptually it is easy to distinguish among these three phases, it is not straightforward to do so from a practical measurement point of view.

The methodological issues surrounding productivity costs are: whether they should be included; the scope of their measurement; whether they are a cost or an effect; and how they should be valued. Several studies have rehearsed these issues (Olsen, 1994; Koopmanschap *et al.*, 1995, 1997; Ratcliffe, 1995; Posnett and Jan, 1996; Brouwer *et al.*, 1997a,b; Johannesson and Karlsson, 1997; Weinstein *et al.*, 1997). Koopmanschap and co-workers (1995) argue that productivity costs are relevant, provided the estimates reflect the real changes in production due to disease. There are also ethical arguments for excluding these costs because their inclusion may appear to favour treatment interventions in persons of productive age (Williams, 1992).

A further source of debate about whether productivity costs should be included centres on the potential inconsistency of adopting a societal perspective within cost-effectiveness analysis. This approach was adopted by Gold *et al.* (1996) and led to a situation where some time costs (those arising from treatment) were measured as costs, and other time (or productivity) costs (those arising from morbidity) were included in the outcome measure. Others have argued that, because the outcome measures in cost-effectiveness/utility analysis are health specific (such as life years gained), the opportunity cost of resources should be defined solely in terms of health, thus implying the exclusion of productivity costs (Gerard and Mooney, 1993).

There is also debate about whether productivity changes associated with morbidity should be treated as a cost (in the numerator of the ratio) or as a health effect (in the denominator of the ratio). Those arguing for the productivity costs associated with morbidity to be included in the denominator of the cost-effectiveness ratio (that is, with health effects), do so in order to avoid double counting (Gold *et al.*, 1996). Double counting may arise if, for example, the impact on patients' leisure time when they are ill is captured in both the productivity costs and the quality of life instrument. This approach is strongly disputed by Brouwer and colleagues (1997a,b), who argue that such an approach misrepresents true productivity costs. The issues surrounding productivity costs is an area of costing where there is little consensus.

**Future non-health service costs** may also be incurred, for example food and shelter costs in years of life added (Gold *et al.*, 1996). In theory, it is the net economic burden of survivors on the rest of the economy that is of interest (i.e. consumption minus production; Gold *et al.*, 1996). This therefore suggests that these costs should be included. It is also possible to argue that unrelated non-health service costs could be omitted because, if these costs were truly unrelated, then their consistent inclusion or exclusion would only add or subtract a constant from the cost-effectiveness ratio (Garber and Phelps, 1997). As with productivity costs, discussed above, future non-health service costs may be excluded if the measure of effect is restricted to health outcomes.

### Non-resource costs

The most common type of non-resource cost is transfer payments. These are flows of money from one group in society to another, such as social security benefits. They are a loss to the payer and a gain to the recipient. They involve no resource consumption, therefore they are usually excluded from economic evaluation, although the administration costs associated with them should be included (Gold *et al.*, 1996). Transfer payments, although not resource costs, may still be a relevant factor from a governmental perspective because government decision makers may want to know the impact of them on the flow of financial resources (Drummond *et al.*, 1997).

## Potential factors influencing the types of cost included

Arguments for including or excluding each specific type of cost in an economic evaluation have been discussed above. These arguments, as well as more general factors, influence the decision on whether these costs should be included. The decision may, therefore, be a combination of several factors and these are now discussed. They are not presented in any order of importance. As will become clear, some investigators argue that economic welfare theory alone should dictate which costs are included and which approach is adopted. Others are willing to adopt more pragmatic positions and prioritise the type of costs included, only collecting information on those costs that are relevant to decision makers, or to prioritise costs in terms of their importance. The factors influencing costs for inclusion are summarised at the end of the section in *Box 7*. Feasibility of measurement should not be a criterion for deciding whether costs should be included (Gold *et al.*, 1996), although ease of measurement clearly has implications for resource-use measurement methods and this is discussed later in this chapter (pp. 13–15).

## Economic welfare theory

Recently published articles have revisited the economic foundations of cost-effectiveness analysis (Garber and Phelps, 1997; Weinstein and Manning, 1997). These followed a concern that economic evaluations, and their guidelines, were being performed or developed in a manner that was inconsistent with economic welfare theory (Birch and Donaldson, 1987; Birch and Gafni, 1992, 1994). These concerns suggest that decisions on which costs should be included should be based on welfare economics alone, and that the concepts of opportunity cost and marginal analysis should dictate what resources are included and how they are measured and valued. The alternative, extra-welfarist position is that the perspective of the decision maker should influence design. If economic welfare theory is the criterion for inclusion, this implies that all costs, apart from non-resource costs, would be included and, consequently, the remaining criteria discussed in this section become irrelevant.

## Perspective

Alternatively, the perspective, or viewpoint, of the analysis influences which costs are included (Drummond *et al.*, 1987; Davidoff and Powe; 1996; Luce and Elixhauser, 1990a). Possible perspectives include the health service and the decision maker, such as the Government, the patient or society. Additional perspectives are the clinician, patient group or purchaser (Drummond and Jefferson, 1996). The adoption of a health service perspective implies that only health service costs would be included; the adoption of a decision-making perspective implies that, for example, future health service costs may be excluded if the decision on whether or not to treat a future condition is a separate decision from the one being addressed (Donaldson, 1990); the adoption of a patient perspective implies that travel costs would be included; and the adoption of a societal perspective implies that productivity costs should be included. Empirical studies, however, vary in terms of their interpretation of the implications of the societal perspective for cost measurement. For example, Phillips and co-workers (1994) stated that they had adopted a societal perspective and included productivity costs, while Mark and colleagues (1995) stated they had adopted a societal perspective but excluded productivity costs. The adoption of a societal perspective implies that all costs, except non-resource costs, should be included, regardless of who incurs them. A societal perspective therefore reduces the likelihood of cost shifting because, for example, the non-health service costs borne by others, such as the patient or social services, are included. For this reason, some authors argue that a societal perspective should be adopted in all cases (Johannesson, 1995; Gold *et al.*, 1996). The Canadian guidelines also recommend a societal perspective (Canadian Coordinating Office for Health Technology Assessment, 1994). Gold and co-workers (1996) argue that a reference case should be used and that this should be from the societal perspective, with any deviations from this perspective explained and justified. Others recommend collecting additional data so that the analysis can be carried out from a number of viewpoints (Canadian Coordinating Office for Health Technology Assessment, 1994; Drummond and Jefferson, 1996).

## Form of economic evaluation

The form of economic evaluation to be conducted, such as cost–consequence or cost-effectiveness analysis, also determines the costs included (although it may not be possible to judge in advance what the most appropriate form of economic evaluation will be; Donaldson *et al.*, 1996). The issue is that there should be consistency between the breadth of the measurement of costs and effects. As discussed under future health service costs (pp. 6–7), if the effects relate to health, as in cost-effectiveness and cost–utility analyses, it has been argued that the costs included should relate to health (Gerard and Mooney, 1993) and thus, for example, productivity costs would not be included. This then relates back to the issue of perspective.

## Double counting

A general consideration in identifying costs for inclusion is to avoid double counting, that is, counting the same cost twice or including an item both as a cost and as an effect (Donaldson and Shackley, 1997). The issue of double counting has already been raised in the context of productivity costs (pp. 8–9), but it applies equally to all types of cost. The potential for double counting is greatest in cost–utility analysis because some consequences may have been included in costs (Johannesson, 1997). For example, a change in leisure time may be captured in the quality of life instrument. The potential for double counting therefore depends on how healthcare and income losses resulting from disease are financed and how the questions assessing quality of life are phrased. If, when asked questions relating to quality of life, respondents are told to ignore the impact that a return to work has on their income, then the morbidity effects on productivity costs would not be double counted (Drummond *et al.*, 1997; Johannesson, 1997).

## Quantitative importance

For particular health service interventions, certain types of costs will be more quantitatively important than others. For example, if evaluating preventive programmes, then future costs may be particularly relevant. Similarly, if evaluating mental health service interventions, costs on other public sectors budgets are likely to be important. In principle, resources that are large enough, either individually or collectively, to have an impact on the cost-effectiveness ratio should be included (Gold *et al.*, 1996) and this should therefore be used as the definition of quantitative importance. This may arise as the result of a small number of resources being consumed by a large number of patients, or a large number of small differences across resource elements may affect the comparison of alternatives. Similarly, cost-generating events with high unit costs may not necessarily dominate the total cost. Cost-generating events with relatively modest unit costs may make an important contribution to the total cost if they occur sufficiently frequently (Spiegelhalter *et al.*, 1996). On p. 7 it was noted that the decision on whether to include future health service and non-health service costs could be based on their expected impact on cost-effectiveness. The quantitative importance of future costs will, however, be reduced by discounting future costs to present values (Donaldson, 1990). Pretrial modelling could be used to establish likely quantitative importance (Sculpher *et al.*, 1997).

If there is no quantitative difference in the magnitude of a type of cost between the arms of a trial, the common resource use could be excluded from measurement because it would not affect the choice between the given interventions. If, however, comparisons of absolute cost are likely to be made with other interventions outside the initial comparison, then excluding common resource use may not be appropriate (Drummond *et al.*, 1997).

The use of quantitative importance as a criterion may be misleading. First, some studies have omitted certain costs, stating that they are unlikely to have an 'appreciable effect' on the cost-effectiveness ratio (Jonsson and Weinstein, 1997), but what can be considered 'appreciable' is often left undefined and is inevitably subjective. Secondly, if the interest is in the quantitative impact on the cost-effectiveness ratio rather than total cost, it ignores the relationship between costs and effects. Small differences in costs are important if there are small differences in effect.

## Attribution

An important design issue is whether to include all resource use or only that related to the disease (Jonsson and Weinstein, 1997). Determining whether or not resource use is related is often a subjective decision and depends on whether its use can be attributed to the disease. The attribution of costs, although particularly relevant for future costs and general illness costs, is relevant for all types of cost.

Attribution depends on at least two factors: first, having an understanding of the reasons why resources are consumed; and, secondly, having clearly defined criteria for attribution. It is not always easy, however, to define attribution according to the underlying clinical reasons for the use of health service resources (Schulman *et al.*, 1996b) and the dividing line between whether or not a cost is attributable is often arbitrary and subjective (Hurley *et al.*, 1995).

Empirical studies have handled attribution differently. In an economic evaluation of anticoagulant therapy after myocardial infarction, costs of fatal events occurring outside the hospital were excluded from the analysis but the reason for this decision was not reported (van Bergen *et al.*, 1995). In an economic evaluation of treatment for menorrhagia, attribution of resource-use events was determined by clinicians (Sculpher *et al.*, 1996a).

Rather than addressing attribution, information on all health service resources used by patients in a trial could be collected, including those resources expended on treating other diseases (Schulman *et al.*, 1996b). The inclusion of all health service resources avoids the neglect of any unexpected resource use that may be causally related to the interventions being compared (Jonsson and Weinstein, 1997), but the risk is that a few high-cost events will ultimately dominate the total cost.

The purpose of costing within a trial is that all confounders have been removed so that differences in cost can be attributed to the intervention and attribution may be resolved by the randomisation procedure in a trial. Schulman and co-workers (1996b) argue that any difference between two treatment arms would be attributable to the study drug if true randomisation were achieved. This would only be the case, however, if the trial had been powered to detect this difference and if the sample size was sufficient for the randomisation process to control for these factors. Other health

service visits have been shown to be highly variable and therefore to require larger overall sample sizes (Wonderling *et al.*, 1996a). This issue is considered again later in this chapter (pp. 15–17). An alternative way of handling the attribution issue would be to adopt the approach used for clinical data, with attribution decided by a panel on a blinded basis. A pragmatic approach to handling the issue of attribution is to present all costs as well as attempting to determine their attribution. Further research on the implications of alternative criteria for attribution is required.

### Time horizon

The time horizon is the period of time for which costs and effects are measured. If a shorter time horizon is used then future costs would not be included. The time horizon of the economic evaluation will be determined partly by the duration of the trial, but also by the perspective of the study and the period of time for which the decision maker has an interest, which may be longer than the trial follow-up period (Davidoff and Powe, 1996). Although the analysis can be performed for a number of time horizons, a long-run perspective has been recommended (Gold *et al.*, 1996). This is because the longer the time period, the greater the number of costs that change or that are variable. Resource use may change over time or new resource items may be consumed, ultimately affecting the direction and magnitude of cost differences.

Limiting the costs of analysis to a fixed period after the intervention may introduce bias into the cost comparison, especially if a disproportionate amount of resource consumption is made near death (Dranove, 1995). This is particularly relevant in advanced cancer clinical trials where, in general, high costs are incurred during the terminal phase of the illness (Bonsel *et al.*, 1993). Sculpher and colleagues (1994) showed that cost advantages change over time and found angioplasty to reduce the initial costs for treatment of angina compared with coronary artery bypass grafting, but that the need for subsequent procedures after angioplasty reduced the balance of cost advantage after 2 years. This result occurred because of differences in retreatment rates. It is possible for the cost gap to change over time because of learning effects (Kesteloot and Penninckx, 1993; Dranove, 1995; Langkilde and Sogaard, 1997). Rather than collecting data over a longer period, an alternative approach is to model results beyond the end-point of the trial. Modelling approaches are discussed in chapter 4 (pp. 32–33).

**BOX 7 Summary of potential factors influencing the costs included**

- economic welfare theory
- perspective
- form of economic evaluation
- double counting
- quantitative importance
- attribution
- time horizon.

## Measuring resource use

### Identifying key cost-generating events

It is not necessary to collect detailed resource-use information for each patient. The key cost-generating events can be measured, for example, a day in hospital or a visit from a social worker. Thus, cost-generating events can occur for both health and non-health service resources. Identifying the cost-generating events requires some prior understanding of the treatments and procedures associated with the interventions. If the key cost-generating events can be determined in advance, then it may be possible to collect data for these events only. These events are likely to be those that have a high unit cost or those with a small unit cost and high frequency.

The advantages of identifying key events in advance are that it limits the data collection effort (Morris and Goddard, 1992; Knapp and Beecham, 1993; Howard *et al.*, 1995), reduces the likelihood of the accuracy of the data being affected (Morris and Goddard, 1992; Drummond, 1994), limits the burden placed on patients (Spiegelhalter *et al.*, 1996), and may reduce research expenditure (Drummond and Stoddart, 1984; Spiegelhalter *et al.*, 1996). The aim is to minimise data collection while maximising the ability to measure the difference in costs (Howard *et al.*, 1995).

The definition of key cost-generating events can be based on the criteria shown in *Box 8* and discussed below.

If the resource-use items that are likely to vary between the arms of a trial can be predetermined, any events common to both arms may be treated as deterministic (O'Brien *et al.*, 1994). Similarly, resource-use data collected during the trial could be for those events that vary unpredictably from patient to patient (Drummond and Stoddart, 1984; Drummond and Davies, 1991; Howard *et al.*, 1995) or between patients within arms. Events such as hospitalisation may involve predictable

**BOX 8 Potential criteria for defining key cost-generating events**

- variation in the frequency of events between arms of the trial
- variation in the frequency of events between patients within arms
- impact on cost/cost-effectiveness ratio
- consequences if data are not collected on the event
- hypotheses about events.

combinations and sequences of resource use and therefore may not require measurement directly from patients (Clark *et al.*, 1994). It has also been suggested that, if items do not contribute a large amount to the total cost, they could be estimated rather than collected (Spiegelhalter *et al.*, 1996). Others argue that resource consumption that is either individually or collectively large enough to have an impact on cost-effectiveness should be included (Gold *et al.*, 1996). Before determining which precise cost-generating events to record, the investigator needs to develop specific hypotheses about resource use (see below, p. 15).

Several methods can be used to determine the key cost-generating events in advance *(Box 9)*. Previous studies may highlight the parameters that are the main determinants of cost (Backhouse *et al.*, 1994). Pretrial data collection may indicate the range of resources to be considered (Drummond and Davies, 1991) or pilot studies could be conducted (Morris and Goddard, 1992; Drummond, 1996; Drummond and Coyle, 1997). In the case of a pharmaceutical trial, Phase II studies should provide valuable information for designing the Phase III studies. Models may be used to highlight the key variables (Sheldon, 1996; Sculpher *et al.*, 1997). Expert opinion could also be used to identify the major resource items expected to be consumed by patients (Schulman *et al.*, 1996b). Claxton and Posnett (1996) suggest that the decision could be based on the marginal value of information provided by each event.

**BOX 9 Possible methods for identifying key cost-generating events**

- reviewing previous studies
- pretrial data collection
- pilot testing
- modelling
- expert opinion
- determining marginal value of information.

The composition of the total cost, and thus the contribution of key cost-generating events to total cost, has been examined empirically. Studies have used detailed cost data to explore the extent to which total cost might have been predicted from collecting data on fewer events (Knapp and Beecham, 1993; Whynes and Walker, 1995). In a mental health context, a list of 21 cost items was reduced to five, which still accounted for 94% of the total cost (Knapp and Beecham, 1993). The authors argued that the use of a reduced list of cost-generating events is appropriate if the interest is in a broad order of magnitude of costs rather than variation in costs between patients. In an acute setting, Whynes and Walker (1995) developed a reduced list of four items from an initial 14, which accounted for 95% of the total cost. They concluded, however, that, because between-patient variation in cost was wide, the reduced-list approach concealed important cost variation between patients. Schulman and co-workers (1996b) decided not to collect frequently performed but low-cost items such as routine blood tests on the basis of data showing that the tests made up only 1.8% of the total procedure costs for these patients.

## Units of resource-use measurement

Resource-use quantities can be defined in different ways. Methods of resource-use data collection are discussed in detail in chapter 3 but, before discussing the methods used, the units of measurement need some consideration.

Health service resource use is usually measured in physical units, such as hours of staff time or doses of a drug. A related aspect is whether to measure the intensity as well as the frequency of resource use (Clark *et al.*, 1994), so, for example, the duration of a GP visit may be of interest as well as the number of GP visits.

A further aspect is whether to measure the change in resource use by measuring the average resource use of interventions A and B or by measuring the increment directly. The former approach is more common but the latter has also been used (Bryan *et al.*, 1995a) and requires an obvious baseline.

The units of resource measurement are also relevant when considering the variability in cost data. If the units of resource-use measurement are detailed, this may lead to higher variability and have implications for sample sizes; this is discussed later in this chapter (pp. 16–17).

## Relating resource-use measurement to valuation

Unit costs are attached to resource quantities to estimate costs. Therefore the definition of appropriate resource quantities must relate to the unit costs available. This will then affect the level of detail at which the unit costs have to be measured. This is true for both health service resource use and non-health service resource use, such as time. Although the purpose of this report is not to review alternative methods of unit cost estimation, some brief points are necessary to address the relationship between resource-use measurement and unit costs. Sources of unit cost data are noted briefly later in this chapter (p. 18).

There is a spectrum of detail (or precision of measurement) of unit cost data (Canadian Coordinating Office for Health Technology Assessment, 1994; Drummond *et al.*, 1997). The ends of the spectrum have been described as gross costing and microcosting (Gold *et al.*, 1996). Gross costing is where resources are identified at an aggregated level and a unit cost is attached. For example, the number of hospital days could be measured and valued by the unit cost of a hospital day. In microcosting, resource use is identified at a detailed level and a unit cost is attached to each resource. For example, staff time spent administering a drug could be measured and valued by staff cost per hour.

It is therefore important to ensure at the design stage that unit costs are available to attach to the chosen measurement of resource quantities. A microcosting approach may require the estimation of separate unit costs for capital and overhead items. Methods for carrying this out are discussed elsewhere (Donaldson and Shackley 1997; Drummond *et al.*, 1997). Similarly, if a resource has been measured at an aggregated level, such as a hospital day, then an appropriate unit cost per hospital day must be available. Unit costs used with gross costing may include *per diem* costs (Donaldson and Shackley, 1997), which incorporate capital and overhead items. The decision on the level of costing (microcosting or gross costing) depends on the availability of unit cost data and the effort to be put into collecting data on unit costs. In some cases, unit cost calculation may require some resource-use information.

There is a similar relationship between the measurement of non-health service costs and their valuation, in that measurement depends on available valuations. Published unit costs of social services are available (Netten and Dennett, 1996). When patients travel by car, costs are measured in miles and valued by a mileage rate. Time costs are estimated by valuing the time by the opportunity cost of forgone activities, such as work or leisure. The methods employed usually distinguish between working and non-working time, with work time valued by wage rate and non-work time valued either by a constant value (Sculpher and Buxton, 1993) or a fraction of the wage rate (Torgerson *et al.*, 1994). Production losses can be measured in terms of days lost and a value attached to this (Drummond *et al.*, 1997). The valuation of production losses is a source of controversy. Traditionally, the human capital method is used, which utilises the wage rate as a measure of lost earnings. The argument for using the wage rate is based on the neoclassical viewpoint that wages equate to marginal productivity, with the assumption of a full rate of employment (Brouwer *et al.*, 1997a). An alternative valuation method suggested is the friction cost method (Koopmanschap *et al.*, 1995; Koopmanschap and Rutten, 1996), which takes labour market conditions, such as unemployment and replacement, into account. The methodological issues surrounding the measurement and valuation of costs incurred by other public sector budgets, such as mental health services, are discussed by Knapp and Beecham (1990). The particular methodological issues involved in the measurement and valuation of informal care are discussed by Smith and Wright (1994). These issues include the potential benefits as perceived by carers. They argue that studies attempting to include informal care costs have used inconsistent methodologies.

## Timing and frequency of resource-use measurement

If the timing of resource-use data collection is addressed at the design stage, the opportunity to coincide it with the collection of clinical data can be maximised (Morris and Goddard, 1992). It is not necessarily the case that data can be collected only when events occur, that is, be resource driven. In many clinical trials, resource-use data can be collected together with quality of life data, when data collection is effects driven. It is important to note that, if using this approach, quality of life data are often assessed on a calendar basis rather than on an event basis (when events occur). Alternatively, it may be more appropriate to time data collection so that it is consistent with the trial period. It may be possible to build upon the trial data collection procedures in order to collect resource-use data (Drummond *et al.*, 1997), thus driving data collection by the trial protocol or

schedule. If this data collection is not coordinated and integrated with trial data collection, it may be carried out in parallel (Mauskopf *et al.*, 1996), but this may increase the burden on the patient in terms of the completion of multiple data forms. There are therefore three possible options for determining the timing of data collection *(Box 10)*.

**BOX 10 Determining the timing of resource data collection**

- resource driven: collect when events occur
- effects driven: collect when quality of life data are collected
- schedule driven: collect when trial protocol events occur.

A further point to consider is the intervals of data collection. With the resource-driven approach, the time cycle of disease determines the intervals. With the other approaches, the time cycle of disease should be considered in addition to the ability of patients to recall information over intervals. Recall is discussed further in chapter 3 (pp. 22–23). Gorsky (1996), in costing an HIV counselling programme, suggested data collection intervals of 1 week. Schulman and co-workers (1996a) collected data biweekly. Time sampling techniques for events that occur frequently can be used and 2-week sampling is often utilised for drug use. Alternatively, respondents could be asked to report resource use for short periods of time that are easy to remember; this information is then used to estimate resource use over a longer period (Clark *et al.*, 1994).

## Sampling strategy

Sampling decisions have to be made in order to test hypotheses and to estimate confidence intervals for cost differences or cost-effectiveness.

### Hypothesis testing or estimation

Whether to adopt an hypothesis-testing or an estimation approach is part of a wider debate in health services research (Gardner and Altman, 1986); thus no correct approach exists. The hypothesis-testing approach aims to test differences between groups in the outcome variable of interest, according to a chosen level of statistical significance. Hypothesis testing places emphasis on the statistical significance of results in isolation from the magnitude of the size of effect (O'Brien *et al.*, 1994). In contrast, an estimation approach focuses on the precision and the magnitude of difference of the estimate. An estimation approach is often preferred because assessing a relevant quantity is more informative than significance value. Confidence intervals indicate the level of precision, or uncertainty, associated with an estimate and convey more information than statistical tests of significance. A potential advantage of confidence intervals is that statistical significance can also be inferred from them; for example, if a 95% confidence interval around a difference includes zero, then the groups compared are not statistically significantly different at the 5% level (O'Brien *et al.*, 1994). Furthermore, some studies may be set up to generate an hypothesis rather than to test one.

### Sample sizes

The calculation of sample sizes to test for differences in clinical outcomes between groups is based on the minimum number of observations required to detect a given predetermined clinically important difference, with a given power, at the conventional level of statistical significance or a prespecified level of precision of estimation. Separate sample size calculations based on detecting differences in economic outcomes are not common and therefore sample sizes are often still based on the clinical outcomes (Drummond, 1994).

Economic outcomes require their own sample size calculations, based on economically important differences (Bonsel *et al.*, 1993), to detect differences between the interventions in terms of cost-effectiveness, costs or resource use. Ideally, sample sizes are calculated to estimate or test for differences in cost-effectiveness but this is technically difficult because it requires that the distinct features of the incremental cost-effectiveness ratio are taken into account. These features are discussed below. Thus, to date, the determination of sample sizes to detect differences in cost-effectiveness has not been fully explored. The steps used for calculating sample sizes for economic outcomes will now be discussed. Those for calculating sample sizes are summarised at the end of this subsection in *Box 11*.

**The first step** is to specify an hypothesis. If an hypothesis-testing approach has been adopted, an hypothesis has to be specified to test for differences in resource use, costs or cost-effectiveness (Coyle, 1996), which also requires specification of an alternative hypothesis. If the important cost-generating events have been established, then sample sizes can be based on detecting differences

in these individual events rather than on detecting differences in total cost (Drummond *et al.*, 1997). Separate sample size calculations are often required for different types of costs, such as patient time costs, and these imply different sample sizes (Drummond and O'Brien, 1993). If sample size calculations are based on testing for differences in cost-effectiveness, then this should take into account the distinct features of the cost-effectiveness ratio. One feature of the cost-effectiveness ratio is that the numerator of the ratio comprises many cost-generating events that have been multiplied by their unit costs. A second feature of the cost-effectiveness ratio is that it is a ratio of two variables with separate and non-independent variances (Mullahy and Manning, 1995; Mullahy, 1996). This requires the use of a formula for determining the variance of the ratio (O'Brien *et al.*, 1994; van Hout *et al.*, 1994). Willan and O'Brien (1996) have demonstrated that it may be possible to use Taylor's expansion to estimate the required sample sizes. At present, the estimation of sample sizes to test for differences in cost-effectiveness is not common.

The sample size calculation also requires selecting a level of statistical significance and the power the study will have in order to detect differences. Gray and colleagues (1997) found that, once the difference has been defined, some clinical trials have, in fact, been too small (or underpowered) to detect the required differences in costs. Other studies have used small sample sizes without discussion of power issues (Drummond *et al.*, 1991).

**The second step** is to determine an economically important difference. This involves specifying the minimum difference in resource use, cost or cost-effectiveness that is considered quantitatively important (Drummond and O'Brien, 1993). There is no consensus on what constitutes this difference. Drummond and O'Brien (1993) suggest that an important cost difference would be any difference greater than the costs of changing to a new method of treatment. Coyle and co-workers (1995) note that, if the difference is an absolute amount rather than a proportion, the economically important difference depends on the magnitude of the total cost. A small cost difference may be more important to interventions with relatively low total costs. Torgerson and colleagues (1995) base the sample size calculation on the difference in effect that would be cost neutral between alternatives. This results in a minimum sample size required to detect a stated difference in effect. If the important difference is seen in terms of cost-effectiveness then the maximum acceptable incremental cost-effectiveness ratio has to be specified. Laupacis and co-workers (1992) suggest a threshold of important difference in cost-effectiveness that would represent good value for money (Canadian $20,000 per quality adjusted life year). This approach was not, however, concerned with statistical considerations and has been criticised for not being based on economic theory (Gafni and Birch, 1993). An economically important difference could be one that is deemed to be policy relevant (Drummond *et al.*, 1997).

**The third step** is to assess existing knowledge on the variability in, and distribution of, resource-use, cost or cost-effectiveness data. The ability to conduct sample size calculations will depend on prior knowledge of the variability of resource use or costs. Data from previous studies could be used to inform sample sizes (Bonsel *et al.*, 1993). If data are unavailable, pilot studies could be performed. The limitations of using pilot studies include restricted follow-up, small sample sizes and low power (Wittes and Brittain, 1990), and these should be recognised when using such an approach. In a study by Schulman and colleagues (1996a), no pilot data were available to calculate the required sample for the secondary economic end-points but, in order to minimise the chance of too little power, multivariate analyses of costs were planned as an alternative to testing for differences between treatments. Drummond and Coyle (1997) simulated a pilot study by analysing data collected over the first 3 months of a study and found that the pilot study identified important cost-generating events but was limited in calculating sample sizes because some knowledge of unit costs was required *a priori*.

The variability (or dispersion) in cost-generating events should also be considered at the design stage because over-dispersion of events may require larger sample sizes (Spiegelhalter *et al.*, 1996). As noted above, variability may depend both on the units of resource measurement used and the number of events measured. If detailed measurement is used and there is high variability, this implies a large sample size. For example, if the costs being measured include all health service costs, then the sample size required to detect differences in total costs (defined as including all health service costs) may be larger than that required to detect differences in costs excluding these costs (Wonderling *et al.*, 1996a).

As a result of high variability in resource use, a larger sample size may be required for the economic outcomes than the clinical outcomes in order to show a difference between groups

(O'Brien *et al.*, 1994; Drummond *et al.*, 1997). This introduces important ethical considerations (Drummond, 1994; Gray *et al.*, 1997). It has been proposed that the level of statistical significance used to test for differences in economic variables may have to be reduced in order to keep the clinical and economic sample sizes the same (Drummond *et al.*, 1997).

Sample size calculations should also take into account the distribution of the data (O'Brien *et al.*, 1994). If raw data are available on which to base a sample size, they should be examined to see if the distribution is normal. If it is not normal then either transformation is necessary before estimating the sample size (Gray *et al.*, 1997) or a non-parametric sample size calculation should be performed. Other forms of sample size calculation may be carried out, such as a formula based on testing the differences between means but controlling for covariates, as used by Schulman and co-workers (1996a).

**BOX 11 Steps in calculating sample sizes**

- set up a hypothesis to test for differences in resource use, costs and cost-effectiveness at a specified significance level and power
- identify an economically important difference
- determine prior knowledge on variability and distribution of cost data from previous studies or from pilot studies.

## Sampling

As previously discussed (pp. 2–3), it may not be necessary to collect all resource-use data from all patients; some resource-use data could be collected from a subsample of patients only. This may be a useful approach for data for which there is no prior knowledge and which is hypothesised not to vary between patients, and in order not to overburden patients with data collection. If it is the case that there is a lack of knowledge and it is not feasible to collect data on the whole sample, then it may be worth while to perform a more precise costing on a small sample in order to provide a more detailed profile of events on a random sample (Spiegelhalter *et al.*, 1996). Data could also be collected on a subsample of patients from outside the trial if there is no reason to expect a difference between the trial and non-trial samples (Howard *et al.*, 1996). It is usually the case that unit costs are collected from outside the trial (Drummond, 1994) (discussed later in this chapter; pp. 18–19). If resource-use data were collected from outside the trial, however, the sample from outside the trial would have to have clinical characteristics comparable with the patients within the trial population. This problem could be lessened if investigators use the trial eligibility criteria to select subsamples. The mechanism for the selection of subsamples is an important issue but it is one that has not been addressed in the literature.

## Centre selection

A number of important issues arise when considering multicentre (and multinational) trial designs. Multicentre studies are often conducted to improve the statistical power of studies so that they can be performed over a shorter time. Multicentre economic evaluations have the advantage of allowing for variations in practice and characteristics, that is, for heterogeneity (Coyle, 1996). The choice of centres for inclusion is often determined by the investigators on the basis of where data are easiest to collect rather than the representativeness of the centre. Centre selection introduces issues of generalisability of results and transferability of results across settings. Issues associated with costing in multicentre trials are summarised in *Box 12*.

**BOX 12 Costing issues in multicentre trials**

- Do centres differ in terms of their economic characteristics?
- If centres differ, should resource-use data be collected at all centres or for a subsample?
- If centres differ, should unit cost data be collected from all centres or for a subsample?

Multicentre clinical trials raise the issue of heterogeneity across centres, which has implications for whether economic data can be pooled. Differences in resource use, unit costs and outcomes may occur and introduce economic bias. This may suggest that it is necessary to randomise patients to centres but this is recognised as being difficult for two reasons (Ellwein and Drummond, 1996). First, sources of bias may not be known in advance and, secondly, it may not be feasible. Although heterogeneity can be addressed at the analysis stage (see chapter 4; pp. 29–30, 32–34), the centres selected for the trial may need to be given more careful consideration if they differ in their economic characteristics. Some studies may have a sufficient number of centres to describe variation in centre characteristics statistically. In selecting centres, a definition of what a representative centre can be taken to be is required (Johnston *et al.*, 1997). A possible approach to centre selection would be to stratify centres by factors that are expected to account for economic differences,

for example, rural/urban or high/low occupancy rates or teaching/non-teaching hospitals.

Disagreement exists about the extent to which both resource-use and unit cost data need to be collected from all centres or countries participating in a multicentre trial. If centres are likely to differ in terms of their economic characteristics, then resource use should be collected from all centres, otherwise the items of resource use that are likely to differ between centres should be collected. If collecting resource-use data from each centre, this has implications for sample size within each centre (Coyle, 1996) and the design of the study. For example, if the unit of randomisation is the centre (cluster randomisation) rather than the patient, this has implications for sample size calculations.

Calculations of unit costs for each centre may not be necessary if detailed resource-use data have been collected (Coyle and Drummond, 1996). The Australian guidelines also argue for a single set of unit cost data (Commonwealth of Australia, 1995). Some empirical studies have used the same unit cost data for all centres, arguing that using centre-specific unit cost data would conceal differences in resource use (Ratcliffe *et al.*, 1996). Centre selection may be a source of bias in unit cost measurement (Jacobs and Baladi, 1996). Unit costs may differ by centre for several reasons, including geographical variation, with, for example, the expectation of higher unit costs in locations in London compared with those elsewhere in the UK (Sculpher *et al.*, 1994). Variation in unit costs between centres and countries may also arise because of differences in relative prices, and opportunities to deploy resources may differ from place to place (Drummond and Jefferson, 1996). Schulman and colleagues (1996b) argue that the unit costs from single centres in a multinational economic evaluation may reflect neither the average unit costs nor the true variation in unit costs and thus unit costs from all centres should be used. Unit costs may also differ if operational efficiency levels and capacity are different (Ellwein and Drummond, 1996). The effect of throughput and capacity on cost and cost-effectiveness can be examined using sensitivity analysis (Sculpher *et al.*, 1992).

Menzin and co-workers (1996) conducted a multinational economic evaluation and collected resource-use data from a single country and unit costs of resources from all the individual countries. In addition, two key resource parameters were identified in the early stages of the evaluation, which were then collected in all countries. In a multinational economic evaluation conducted by Schulman and colleagues (1996a), unit costs of resource use were collected from one centre only, but differences across centres were addressed by generalising the results using sensitivity analysis. Drummond and co-workers (1992) found that cost-effectiveness was affected by practice patterns and payment systems, and identified differences across centres in terms of resource use and relative prices. In examining centre differences, a recognition of the potential relationship between resource use and unit costs is required because, if unit costs are a function of resource use at individual centres, this implies that centre-specific unit costs should be used.

## Data required from outside the trial

Data other than that derived from the trial is required in order to perform the economic evaluation (Powe and Griffiths, 1995; Rittenhouse, 1995, 1997). There are several reasons for this, which are now discussed.

### Unit cost data

In practice, unit costs of health service resources are derived from outside the trial from a variety of sources. If a microcosting exercise is performed, unit costs are most likely to be derived from the local hospital finance departments. For gross costing approaches, published costs in previous economic evaluations or other published cost data may be used (British Medical Association/Royal Pharmaceutical Society of Great Britain, 1996; Office of Health Economics, 1997a; Netten and Dennett, 1996). Sources of unit cost data in the UK differ from other countries, particularly the USA, where charge data are readily available and valuation issues focus on cost to charge ratios (Copley-Merriman and Lair, 1994). Discrepancies between charges and cost data have been examined elsewhere (Cohen *et al.*, 1993). In the UK, most hospitals have extracontractual referral tariffs, which might be seen as corresponding to American charges, but may not reflect the costs of the bulk of patients covered under large contracts. There is also a growing database of costs per health care resource group (CHKS, 1996). In the UK, there is likely to be a growing availability of cost data with the development of the reference cost schedules by the Department of Health.

The quality of the unit cost data is important and biases in unit cost measurement may exist (Jacobs

and Bachnynsky, 1996) in addition to centre selection bias (noted above, p. 18). There are at least three additional sources of bias in unit cost measurement, but these could be adjusted for by using sensitivity analysis (Jacobs and Baladi, 1996). First, a scale bias may exist if it is assumed that the marginal and average costs are the same. The difference in these two estimates is discussed by Goddard and Hutton (1991). Secondly, there may be a methods bias if the methods used lead to different cost estimates. Thirdly, a casemix bias may exist if costs are based on a particular type of patient only. Furthermore, unit costs should theoretically reflect opportunity costs, although they often do not. Some studies have, however, explored the valuation of staff resources in terms of opportunity cost (Hughes *et al.*, 1996; Ratcliffe *et al.*, 1996). They have done this noting that an opportunity cost is only incurred if staff time released is used in an alternative way. Staff costs can be valued at zero opportunity cost in a sensitivity analysis if no alternative use exists.

### Data required to generalise results

A commonly cited limitation of clinical trials is the low external validity of the results, that is, the low level of generalisability to other populations and settings (Bailey, 1994; Davis, 1994; Rubins, 1994; Fayers and Hand, 1997). This usually refers to the distinction between efficacy and effectiveness (Simon *et al.*, 1995) but it applies equally to cost data. Factors limiting external validity include the fact that the clinical outcomes realised in a non-trial setting may be different owing to different locations, differences between the trial protocol and routine practice, and atypical patient compliance within clinical trials (Drummond and Jefferson, 1996). It is important, therefore, to generalise from the trial to usual practice, or other settings, for policy-making purposes (Glick *et al.*, 1988; Neumann and Johannesson, 1994; Bryan and Brown, 1997; Mason, 1997; Phelps, 1997).

At the design stage, the extent to which external validity can be maximised should be considered; this may involve collecting additional data. For example, information on normal clinical practice at the trial site could be collected so that the user of the study could translate results to their own circumstances (Drummond *et al.*, 1996). Rittenhouse (1997) proposes using sensitivity analysis or modelling as a way of making adjustments to data collected from the trial situation. It may be possible to generalise using data from other studies, but this requires the consistent reporting of data (Rigby *et al.*, 1996).

### Data required to model long-term outcomes

Clinical trials are usually conducted for a limited period of time, yet the real interest may be in examining the costs and benefits over a longer period (Sheldon, 1996). Trials often employ intermediate health outcomes, whereas economists want to focus on 'final' outcomes (O'Brien, 1996). Consequently, some form of modelling may be required to extrapolate beyond the end-point of the clinical trial and to adjust or supplement the original data (Drummond and Jefferson, 1996). Sheldon (1996) argues that extrapolation is not a reliable substitute for longer follow-up, but it is also argued that failure to use models to extrapolate results may lead to greater errors (Gold *et al.*, 1996). Modelling can therefore be seen as a pragmatic response to the problem of limited time horizons. Data will therefore be required to populate the model. Forms of modelling are addressed in chapter 4 (pp. 32–33).

## Analysis plan

An analysis plan for costing is useful in finalising study design. This should include details of the descriptive and multivariate analyses planned and the modelling approach, if any, to be adopted (Mauskopf *et al.*, 1996). They should also explain how resource-use and unit cost data are to be synthesised (Coyle *et al.*, 1995) and how missing data are to be handled. It may not, however, be possible to predetermine the precise form of economic analysis to be conducted in the analysis plan (Donaldson *et al.*, 1996) because the final form of analysis is conditional on whether the clinical effect is the same, better or worse (Drummond *et al.*, 1997). The criteria used to choose between the forms of economic evaluation should, however, be stated in the analysis plan.

## Summary

Design issues address the types of cost to be included, such as health service costs, trial costs, future costs and productivity costs. The decision about which types of cost to include depends on the seven key factors already identified:

- possible links to economic welfare theory
- the perspective to be adopted
- the form of economic evaluation
- avoidance of double counting
- the quantitative importance of the type of cost

- whether the cost can be attributed to the intervention
- the time horizon of the study.

The collection of detailed data on resource use for all patients may not be necessary and key cost-generating events can be measured. These can be defined as where there is variation in the frequency of events between the arms of the trial or between patients within arms. Determining sample sizes for detecting differences in costs or cost-effectiveness involves identifying an economically important difference and having information on the variability of cost data from previous studies or from pilot studies. A further sampling issue to be addressed in multicentre trials is the selection of centres and whether resource-use and unit cost information should be collected from all centres.

The main design issues where there is a lack of consensus include the measurement of productivity costs and whether future costs, particularly future health service costs in life years gained, should be included. Attribution is an important consideration for including costs, but the implications of alternative criteria for attribution need to be further developed. Sampling issues need additional research, including the issue of centre selection and determining sample sizes for cost-effectiveness.

# Chapter 3

# Data collection

In this chapter, a number of issues relating to alternative data collection methods are discussed. Data can be collected on either a patient-specific or a non-patient-specific basis. Modes of collection include:

- patient self-report (by interview, questionnaire, case record forms, diary cards, standardised instruments)
- health records (including patient notes)
- administrative databases
- expert panels.

The data collection methods adopted may differ according to the types of resource-use information being collected, such as health service resource use or patient time costs. All modes of data collection can be used with prospective designs, where data are collected when an event occurs, and with retrospective designs, where data are collected after the event has occurred. The method of data collection adopted will also affect research costs and response rates.

## Patient-specific resource-use data from patients

### Interviews, questionnaires and case record forms

**Interviews** may be either face to face or conducted by telephone. The latter method has been used to collect resource-use data (Mark *et al.*, 1995). Anie and co-workers (1996) describe the development of a computer-assisted telephone interview to collect morbidity and drug-use information from asthma patients. They tested the validity of this interview against face to face interviews and concluded that it can provide a repeatable and efficient measurement of drug use and health. In general, authors do not report or test the validity of the methods used.

**Questionnaires** may be administered by post or given to patients to complete when they attend for their appointments. Postal questionnaires may be a cheaper method of data collection compared with interviews (Streiner and Norman, 1989). Traditionally, a low response rate has been the major drawback of this method, although further research suggests that improved response rates of over 75% are possible with a general mailing and 90% to a targeted group, such as GPs (Streiner and Norman, 1989). Sculpher and colleagues (1996a) administered a postal questionnaire to assess resource use by women who were treated for menorrhagia; they obtained a response rate of 79%.

Patients' travel costs are often collected by questionnaire (Cook *et al.*, 1994; Bryan *et al.*, 1995b), as is the measurement of patient time costs (Sculpher and Buxton, 1993). There is no standardised questionnaire available for the collection of travel and time costs and there is a tendency to develop new questionnaires for new studies. For productivity costs, van Roijen and co-workers (1996) have developed the health and labour questionnaire. This enables the collection of quantitative data on the relationship between illness, its treatment and work performance by examining absence from paid work, reduced productivity at work and unpaid labour, such as child care.

**Case record forms** are often used in clinical trials to collect clinical data. They were used to collect resource-use data by Mark and colleagues (1994). Case record forms refer to any form asking or recording clinical information but they may be adapted to include both resource-use and quality of life questions. In a study designed by Glasziou and co-workers (1997), resource-use questions were added at the end of the quality of life instrument. Adding questions into case record forms results in a longer form and, potentially, if too many questions are included, this may reduce completion rates as a result of overburdening the patient.

For all three methods, the length of time patients take to complete the interview, questionnaire or form is also an important consideration, particularly when the research involves measures not only of resource use but also of quality of life from the same patients. Indicators of patient burden include refusal rates, rates of missing responses and complaints.

### Diary cards

Diary cards can be used to collect patient-specific data (Fayers, 1995). For example, patients may be asked to record resource use in a diary at home,

either when care is received or on a daily or weekly basis (Mauskopf *et al.*, 1996). Two important considerations when using diary cards are whether patients fill them in at the time of receiving care or whether they wait to fill them in immediately prior to the return date. These design aspects will affect the validity of the method. Verbrugge (1980) conducted a literature review of the use of health diaries with particular reference to their reliability, efficiency, validity and cost. He concluded that diaries produced more accurate content and better data quality than questionnaires. A disadvantage of using diary cards is that patients may fail to complete them, although this could be overcome through reminder telephone calls (Mauskopf *et al.*, 1996). The costs of data collection by diary cards may therefore increase if researchers are required to motivate patients to complete them and to monitor their completion. Thus, diary cards may also have additional data entry and analysis requirements. Patient cooperation with diary filling has been found to be good, although sensitisation may occur initially, when completing a diary stimulates patients to take more interest in their condition, thereby recording more resource use (Verbrugge, 1980). Patient fatigue may also occur and it has been shown that, in studies conducted over 1–2 months, a reduction in motivation may occur and hence the thoroughness of the reporting is reduced (Verbrugge, 1980). Diary cards may provide more accurate data on some patients but a potential drawback could be missing data for others (Weinberger *et al.*, 1993).

## Standardised instruments

Many investigators design a data collection instrument for each study, resulting in wide variations in both the information collected and its accuracy. These instruments or adaptations are rarely tested for validity and reliability. Although it can be argued that the use of a flexible data collection instrument can improve the validity of the data (Clark *et al.*, 1994), flexibility limits the ability to compare the results across studies. In response to this, some researchers have developed standard instruments for collecting resource-use data. An American modular instrument, the Resource Utilisation Survey, has been developed for use in clinical trials (Copley-Merriman *et al.,* 1992). The Personal Social Services Research Unit at Canterbury (UK) have developed the Client Service Receipt Interview, which is designed to gather retrospective information on health and social services resource use in a community/mental health context (Beecham, 1994). The Client Service Receipt Interview has been used in empirical studies (Knapp *et al.*, 1994). As noted above, van Roijen and colleagues (1996) have developed a questionnaire to collect data on productivity costs.

Standardised instruments may improve the ability to replicate and compare results across studies. If service provision has changed since the instrument was developed, the instrument may not be able to detect these changes or the associated changes in resource use over time (Clark *et al.*, 1994). If standardised instruments are to achieve the aim of comparability of results, then investigators should ensure that they do not adapt them.

## Sources of bias in measuring patient-specific resource use from patients

Four potential sources of bias exist when deriving data from patients: recall, non-response, evasive answer, and selection biases. The question format may also affect replies. Bias may be in either direction.

### *Recall bias*

For all patient-specific resource-use events, the selection of the recall interval is important in order to minimise recall bias. This occurs when respondents are asked to report past events based on memory. Recall bias has been addressed in several studies (Brown and Adams, 1992; Revicki *et al.*, 1994). Brown and Adams (1992) reviewed studies looking at hospitalisation, ambulatory care, treatments given and drugs prescribed, and concluded that recall did not appear to deteriorate over a 2–3-month period. Weissman and co-workers (1996) concluded that self-report information from AIDS patients provided valid data with a recall of 4 months or less. Revicki and colleagues (1994) tested the accuracy of recall at 1-month and 3-month intervals for days off work due to disability. They found that an individual's reports were as accurate for 3-month recall as for 1-month. Seasonal variation should also be taken into account when selecting the recall interval (Gorsky, 1996).

Although there is no conclusive evidence regarding an appropriate recall interval, there is evidence that explains the factors influencing recall. The reliability of recall depends on the level of use of health service resources, the severity of the illness, the frequency of occurrence of resource-use events, the conditions under which the interview takes place, and the type of resource use. These are now discussed.

Accurate recall may be limited if the patient is a high user of medical services or if illness interferes with the patient's mental status (Mauskopf *et al.*, 1996). The more severely ill the patient, the worse

his or her recall will be. Cleary and Jette (1984) investigated whether patient characteristics were related to error in reporting and found patients' reporting errors to be higher for those using the health service more intensively.

Recall may also depend on the frequency of events. Jobe and colleagues (1990) found that frequent events were under-reported by 20% or more when respondents were asked retrospectively over a 6-month period. Respondents may use different cognitive processes when recalling frequent events than when remembering less common events (Blair and Burton, 1987). If interviews are used, the length of the recall period is important. Some authors argue that it is best to make the recall period as short as possible (Clark *et al.*, 1994). The issue of frequency of events and time sampling has been discussed in chapter 2 (pp. 14–15).

Recall may also be influenced by the conditions of the interview. It is improved by face to face interviews and is also better for open-ended than for closed questions (Dex, 1991). Recall may also depend on the type of resource use. Paganini-Hill and Ross (1982) found that patients could recall their medical and drug usage but that the agreement between data from interviews and medical records varied by type of drug.

#### *Evasive answer bias*

Evasive answer bias may occur because respondents may wish to hide events (Clark *et al.*, 1994). If answers to interview questions are embarrassing, responses may not be forthcoming or truthful. Evasive answer bias may therefore lead to partial non-response (i.e. respondents may answer some questions only, leading to missing data). Rittenhouse (1996a,c) has reported the development of the 'randomised response interview' technique. This elicits information on the proportion of the population engaging in embarrassing behaviour. Rittenhouse (1996a) argues that randomised response interviews can also be used to derive patient-specific resource-use data and could be utilised where information on medication non-compliance is sought.

#### *Non-response bias*

Non-response bias may arise for many complex reasons. It may occur in all types of patient self-report and can be either full or partial. Partial non-response, as discussed above, is when respondents answer some questions only. Hebert and co-workers (1996), cautioned against using postal questionnaires in patient populations aged over 75 years because non-respondents in this age group were found to be more cognitively impaired, more disabled and had a higher 1-year mortality rate. They found that a high proportion of individuals (22% of 187) refused to participate in home interviews but that the characteristics of non-responders were, apparently, no different from responders.

Poor response rates may introduce bias and limit generalisability (Leigh Brown *et al.*, 1997). The rate may be improved by sending out reminders. Response rates from postal questionnaires may also be improved by offering incentives for individuals to complete them, such as payments or prize draws. Leigh Brown and colleagues (1997) performed a randomised trial on the effect of a prize draw on response to postal questionnaires and found that this produced an extra response, but that the results were inconclusive. In deciding whether to use incentives, the possibility that individuals may find the offer of an incentive offensive may reduce the likelihood of completion and thus have the opposite effect to that intended.

#### *Selection bias*

Selection bias may occur if the population providing the data is not representative of the population at large. The appropriateness of patient self-report methods may depend on the severity of the patient's illness. Weissman and co-workers (1996) suggest that the sickest patients may be unable to participate.

#### *Question format*

The design of the questions may also affect results. Question format is therefore important in order to maximise the quality of the data (Wright and Lawrence, 1994). Several authors have discussed how to phrase questions to minimise any bias (Sudman and Bradburn, 1982; Stone, 1993). Questions should be phrased in an unambiguous manner, and be brief and precise (Stone, 1993).

## Patient-specific resource-use data from questionnaires and case record forms completed by others

Questionnaires or case record forms designed to collect resource-use information may be completed by others, such as hospital staff (Hundley *et al.*, 1995; Hughes *et al.*, 1996) or research nurses. If this approach is adopted, then guidelines on how to complete the forms should be provided in order to maximise the quality of the data (Mauskopf *et al.*, 1996). One limitation of using others to

complete forms is that they may forget to do so, particularly if they are working in busy clinics. Computers could be used to prompt recording activity (Gorsky, 1996).

## Patient-specific resource-use data from existing records and databases

It is sometimes possible to derive patient-specific resource-use information from routine health service records, such as patients' notes, hospital records, GP records or laboratory records. Computer linkage systems are a further type of secondary data set.

### Health records

Health records include patients' notes and other clinical records such as surgeons' notes. If detailed resource-use information is required, records may be a better source of data than burdening patients with a detailed data collection procedure (Mauskopf *et al.*, 1996). Even when records exist, they may be incomplete and difficult to access or retrieve (Clark *et al.*, 1994). They may also require the patient's consent, which may not be forthcoming.

There is no gold standard measurement; many studies have therefore compared patient self-report methods with health records (Linet *et al.*, 1989; Brown and Adams, 1992; Boyer *et al.*, 1995). Linet and colleagues (1989) reported substantial agreement between patient self-report and medical records for conditions and surgical procedures such as splenectomy and appendicectomy, but poor agreement for chronic bronchitis, psoriasis and most types of allergy. Among self-respondents, agreement was greater between males than females, Caucasians than non-Caucasians, and patients who were referred from hospital than those referred from the community. The authors suggested that medical record data collection may be needed to supplement data from interviews for certain conditions.

Harlow and Linet (1989), in a review of the literature on studies comparing data from medical records and patient questionnaires, found good agreement between the two methods for prior hospitalisation and surgery, but poor agreement for X-ray exposure and variable agreement for diagnosis or previous drug use. Agreement depended on the type of resource use.

As with data collected directly from patients, the type of resource use may affect the validity of the method. Additional factors influencing the validity of health records data are the clinical competence of the recorder, the type of provider and setting of care, and the coding and type of data.

### Administrative databases

Administrative databases may be used to examine resource utilisation (Lave *et al.*, 1994). In the USA, these databases relate to clinical and claims databases, which are less common in the UK. The advantage of administrative databases is that they are inexpensive to use compared with primary data collection (Mullahy and Manning, 1995). There are several factors to be addressed when considering using administrative databases to derive patient-specific resource use. First, missing data might be a problem because these databases contain information that is useful only for re-imbursement and may not have been designed to record all relevant information for research purposes. Secondly, the extraction of health record data is subject to problems relating to the level of coder training and the amount of interpretation necessary (Aaronson and Burman, 1994). Thirdly, measurement error may occur because there is a risk that the data in such a database could be an artefact of the way in which the data were collected. Misclassifications in diagnoses can occur, leading to substantial biases (Safran, 1991). The accuracy of the data may also be related to the intended purpose of the database (Wolff and Helminiak, 1996). Fourthly, differences between patients may reflect measurement error rather than any real differences and individual systems may have their own reporting practices (Wolff and Helminiak, 1996). This arises because, although resource use may be identifiable, the reason for resource use and associated diagnoses may not be available. Finally, records may become dated and irrelevant with the passage of time.

### Computerised record linkage

Computerised record linkage may be a useful method of data collection, although this is not widely available. Several such systems have been established in the UK (e.g. the Scottish Record Linkage System and the Oxford Record Linkage Study). The validity of the Scottish system was tested against patient follow-up in a primary prevention study for coronary heart disease. Adverse events were recorded at follow-up trial visits as well as by linking the computer records of all deaths, cancers and hospitalisations to the Scottish database by name, date of birth and postcode. Results showed that the identification of follow-up information based on computer linkage

alone was as effective as patient follow-up. The method was recommended as an accurate and cheap follow-up method (West of Scotland Coronary Prevention Study Group, 1995). The decision to use such a system depends on its cost. If a linkage system already exists then it may be worth while exploring the possibility of using it, otherwise the costs of setting up a system may be prohibitive.

## Non-patient-specific resource-use data

Where a cost-generating event is not considered to be a key factor, it may be sufficient to measure it on a non-patient-specific basis, which is now discussed.

### Existing records and databases

Existing records and databases may be used to derive non-patient-specific resource-use data. For example, length of stay data may be derived from a healthcare resource group or from other routine data. The general limitations of administrative databases have been discussed above (p. 24). Their use as estimates of non-patient-specific resource use raises further issues. Although administrative databases have the advantage that they are based on large samples (Mullahy and Manning, 1995), sample selection may occur; that is, the population coverage of the database may not be representative of the population at large (Kuykendall and Johnson, 1995). The casemix of patients in the databases should have the same casemix as the patients in the trial (Howard *et al.*, 1995).

### Observational techniques

Non-patient-specific resource-use information may be collected by observational techniques, such as time and motion studies or the programme evaluation research technique. These methods are often used to estimate total workload in order to calculate costs (Mugford, 1995).

Time and motion studies require the investigator to produce a process flowchart that includes all steps, for example, in drug preparation and administration. If some tasks are performed simultaneously then this must be accounted for, otherwise the costs will be overestimated (Dranove, 1995). Patient flow analysis may also be used (Gorsky, 1996). This requires measuring the time from when the patient arrives to being seen by different healthcare workers. This may be a useful method for screening or clinic visits when the patient may see more than one member of staff.

The programme evaluation research technique has been used to measure staff time for procedures (Hurley *et al.*, 1994). It involves recording the minimum, maximum and most common time required to complete a task and then estimates the expected time for a task.

Other studies use the observation of a few patients as the basis for their resource-use estimates. For example, Flannelly and co-workers (1997) performed observations at three health centres in order to estimate the costs of cervical screening at GP practices.

### Expert opinion

Estimates of resource use are sometimes derived by ascertaining expert opinion by using consensus gathering techniques. Although this method is unlikely to form the main mode of data collection for patient-specific resource use in a trial, it may be a useful way of estimating non-patient resource use for rare events. There are several methods of obtaining expert opinion, such as structured interviews with expert panels and Delphi panels. Evans (1997) reviewed consensus gathering methods and found that important aspects of designing panel studies included the criteria for selecting experts and the baseline information provided to the panel. Expert panels have been used in several studies (Buxton *et al.*, 1991; Bloom *et al.*, 1993; Phillips *et al.*, 1994; Jonsson and Bebbington, 1994). The Delphi method is a systematic process for eliciting subjective opinion in order to derive a numerical estimate for the variable of interest (Brook *et al.*, 1986; O'Brien *et al.*, 1995).

Expert panels are most commonly used in decision analytical models, but they could be used in clinical trials-based studies. Expert opinion can be used in conjunction with a Bayesian approach to elicit prior beliefs and knowledge about resource use (Jones, 1995). Bayesian statistics uses prior information on beliefs as the basis of estimates. The advantage of expert panels is that they may be an inexpensive, cheap, quick and easy method of establishing resource use, particularly in economic evaluations that are performed after the clinical trial has ended. The limitations of basing resource use on panel estimates is that it may be an inaccurate method if recollection is poor and if estimates relate to ideal service rather than what happens in practice (Drummond, 1994; Drummond and Jefferson, 1996).

## Mixed data collection methods

Mixed modes are often used to collect data on different type of resource use. Patient self-report and records are often used together, either sequentially, in a two-step process, or concurrently (Mauskopf *et al.*, 1996). An example of a two-step process is where patients report the number of events and investigators follow up on the detail of those events through records. Since the problem of defining a gold standard method of data collection remains, multiple sources of data could be used for the same information in order to test for validity (Clark *et al.*, 1994).

## Organisation of data collection

The piloting of data collection methods is an important design aspect. If a Phase III pharmaceutical trial is being designed, piloting could be performed in Phase II (Mauskopf *et al.*, 1996) or conducted in a separate exercise. Activities such as data monitoring are important organisational aspects (Coyle *et al.*, 1995; Mauskopf *et al.*, 1996). Steps can also be taken to assess quality assurance, such as visits to each centre and maintaining contact between site visits (Morris *et al.*, 1993).

Further organisational aspects relate to quality control issues, such as the training of staff involved in the trial and having clear procedures for data entry.

## Summary

Data collection issues involve deciding on the appropriate resource-use data collection method. Resource-use data can be measured on a patient-specific basis by using, for example, interviews, questionnaires, case record forms or diary cards. In selecting a method, potential sources of bias have to be addressed, including recall bias, evasive answer bias, non-response bias, selection bias and question format. If existing health records are to be used, then consideration must be given to whether they will be easy to access and retrieve and the likelihood of patient consent being forthcoming.

The methodological issues in data collection when there is a lack of consensus are the relative performance of alternative data collection instruments and, in particular, the validity and reliability of these instruments. In general, authors do not report or test the validity and reliability of the methods used.

# Chapter 4
# Data analysis

## Summarising cost data

### Constructing total cost

Once resource-use and unit cost data have been collected, a cost per patient can be calculated. For each resource item, the resource quantity used by each patient is multiplied by the unit cost of the resource item. These costs are then summed for each intervention, or arm of the trial, to give the total cost per intervention, or per arm of the trial. At its simplest, the mean total cost for each intervention or arm of the trial is calculated by dividing the total cost by the number of patients in each arm. Certain features, such as missing data, complicate this process and need to be addressed (pp. 27, 33–34). Statistical analysis of cost differences between arms can then be performed as discussed on p. 28.

Alternative approaches for constructing total cost are sometimes used. Some empirical studies have applied unit costs to only those resource items for which there was a statistically significant difference by arm of the trial (for example, Hundley *et al.,* 1995). This approach is also argued by Jonsson and Weinstein (1997). The omission of a resource item that differs between groups should be avoided if the resource item is correlated with other resource items that comprise total cost (Coyle, 1996). This is discussed further below (p. 29). An alternative approach is to estimate total cost by applying unit costs to the average level of resource use per resource item. This is sometimes used to overcome problems of missing data but it does not make maximum use of the patient-specific nature of the data.

The mean total costs are then presented for a relevant period. For example, costs may be shown as weekly (Gray *et al.*, 1997), monthly (Gruger and Backhouse, 1997) or annual amounts (Rutten-van Molken *et al.*, 1995). The period for which costs are calculated depends on the duration of the trial, the period to which the data collection relates, and the length of trial follow-up. Cumulative costs can also be calculated for any given period of follow-up (Hlatky *et al.*, 1997a), for example, at the end-point of the study or as life-time costs. Calculating cumulative costs may require handling censored data and this is discussed later in this chapter (pp. 33–34). Total costs can be presented as a ratio of control group to study group costs, with a ratio greater than 1 indicating higher costs in the control group (Burns *et al.*, 1993; Gray *et al.*, 1997).

Costs are generally measured in a base year, that is, adjusted for the effects of inflation. Discounting of costs and effects is then usually performed, which converts future costs (and benefits) into present values, thereby allowing comparisons between costs and benefits that occur at different times (Drummond *et al.*, 1987; Luce and Elixhauser, 1990a). The differential timing of costs between the study and control arms of the trial can be accounted for by discounting according to the time elapsed since randomisation (Sculpher *et al.*, 1994). Recommendations concerning the appropriate rate of discount vary, as do the discount rates used in practice. Recommended rates of discounting are: 6% (real) in the UK (Her Majesty's Treasury, 1997); and 3% (real) in the USA (Gold *et al.*, 1996), with testing of alternative rates to 5% and 7% in sensitivity analyses (Gold *et al.*, 1996). Whether, and at what rate, health benefits should be discounted is an area of methodological debate (Cairns, 1992; Parsonage and Neuberger, 1992; Krahn and Gafni, 1993) and is not discussed here.

### Issues to be addressed in handling cost data

Certain features of the cost data may have to be addressed before they can be easily summarised. These include whether it is appropriate to pool data from centres, and whether there is missing or censored data. Possible approaches to handling these issues are discussed below (pp. 33–34).

In multicentre clinical trials, resource-use and/or unit cost data may have been collected from multiple centres. Differences among centres (heterogeneity) may exist and be considerable (Bonsel *et al.*, 1993; Mullahy and Manning, 1995). Heterogeneity in resource use may arise because of different practices and policies, while heterogeneity in unit cost data may be a result of different cost structures at individual centres (Bonsel *et al.*, 1993).

Missing data is a methodological problem that is common to many studies in the social sciences (Little and Rubin, 1987). It is important to know why data are missing and what implications the missing data may have (Mullahy and Manning,

1995). Data may be missing because patients did not return or complete the data collection instruments. They are seldom missing at random and therefore the analytical method used in the presence of missing data has implications for the results (Crawford *et al.*, 1995).

Missing data may also arise because of loss to follow-up; these are termed censored data. If patients have not been followed for the entire duration of the trial, then their costs are unknown. Censoring of cost data occurs: when there is loss to follow-up; when a patient drops out of the trial; and because of sequential admission, when patients enter a trial at different times and thus have variable periods of follow-up (Fenn *et al.*, 1995, 1996). Right censoring occurs if patients are not observed at the end of the trial; loss to follow-up is an example of right censoring (Etzioni *et al.*, 1996). Left censoring occurs if patients are not observed for a period at the beginning of the trial. Right censoring is therefore more common. The nature of the censoring may have implications for the methods of analysis in the presence of such data and this is discussed later in this chapter (pp. 33–34).

### Synthesising costs and effects/benefits

In cost-effectiveness and cost–utility analysis, results are usually presented as incremental ratios; in cost–benefit analysis, they are presented as a net benefit/cost. Higher cost-effectiveness ratios indicate lower cost-effectiveness. Incremental cost-effectiveness ratios are calculated only for interventions that are more costly and produce greater effects than the intervention being compared. If more than one strategy is being compared, they can be ranked in terms of increasing effectiveness and dominated strategies omitted from further consideration. Strategies or interventions are dominated if an intervention is less effective and more costly than an alternative; if an intervention is less costly and more effective then the treatment dominates (Karlsson and Johannesson, 1996). This so-called 'win–win' situation was the case for van Bergen and co-workers (1995) in a study of long-term anticoagulant treatment after myocardial infarction. In such cases, no incremental cost-effectiveness ratio needs to be calculated. Average cost-effectiveness ratios, which are created by dividing study arm costs by study arm effects and by dividing control group costs by control group effects, and then taking the difference between the two, are not recommended because, usually, a newer intervention is being compared with an existing one and therefore the most appropriate cost-effectiveness information is extra (or incremental) cost-effectiveness (Briggs and Fenn, 1997). If an intervention is more costly, then an additional judgement may be needed about whether this extra cost is worth incurring, because resources have to be obtained from elsewhere in order to introduce the intervention. An important issue where there is a need for further empirical studies is to explore how costs can best be related to effect, end-points and quality of life.

## Handling uncertainty

The examination of uncertainty surrounding costing data may be considered in terms of resources, total costs and cost-effectiveness.

### Sources of uncertainty

Uncertainty may arise from a number of sources. These include the data inputs, methods and assumptions, as well as the extrapolation or generalisability of results (Briggs *et al.*, 1994). Uncertainty may exist with respect to the parameters used as data inputs, such as the unit costs or discount rate. Uncertainty may relate to the methods used, such as the data collection methods or model, as well as to the assumptions made by the investigator. Uncertainty in the assumptions of the analysis may be greater than any sampling error (Canadian Coordinating Office for Health Technology Assessment, 1994). The extrapolation process itself is a further source of uncertainty and should be subject to testing by sensitivity analysis (Drummond and Jefferson, 1996). The generalisability of results to other settings is a further source of uncertainty.

There are two main reasons for quantifying uncertainty: first, to have some indication of the level of confidence to be placed in the results; and secondly, to test hypotheses about the direction and magnitude of the cost-effectiveness ratio (Gold *et al.*, 1996). The two main methods for analysing uncertainty are statistical analysis and sensitivity analysis, discussions of which now follow. The sources of uncertainty described above use the term 'uncertainty' in a very wide sense, but the methods of handling uncertainty outlined below tend to address a more limited set of issues.

### Statistical analysis

The advantage of basing an economic evaluation on trial data is that the sampled data make it possible to use standard statistical methods, both to examine uncertainty and to test for differences between groups (O'Brien *et al.*, 1994; Spiegelhalter *et al.*, 1996). Whether to examine uncertainty using confidence intervals or perform statistical tests of difference and significance relates to whether the interest is an estimation or hypothesis testing, as

discussed in chapter 2 (p. 15). A further type of statistical analysis that is available to examine uncertainty concerns Bayesian methods (Jones, 1995; Gold *et al.*, 1996) but these are not discussed in this report.

### *Distribution of cost data*

The examination of the distribution of cost data is important because a relatively small number of high-cost events may skew the data. One way of examining the distribution of data is by fitting a probability density function to a histogram of costs (Jones, 1995). Rutten-van Molken and colleagues (1995) examined the distribution of costs in their study and found that, because of a few patients who were contributing large costs, the distribution was heavily skewed to the right. The implications of skewed data are that the mean cost and the variance about the mean are disproportionately affected and the ability to detect significant differences between patient groups is perhaps reduced (Drummond and Davies, 1991). The variability in cost data may be greater than for clinical data and this may result in the trial being underpowered to detect differences in mean cost (Drummond, 1994). A further implication of skewed data is that parametric tests may no longer be used, unless the data are transformed. Some authors argue, however, that, as sample size increases, the asymptotic normality of the mean may approach a normal distribution and therefore parametric tests remain useful (van Hout *et al.*, 1994). Non-parametric tests can be used if no transformation is made. Methods used to compare costs between groups are discussed below (see 'Uncertainty in resource use and cost').

Transformations can be used to reduce the impact of high-cost events on the mean and the appropriate transformation to use depends on the shape of the curve. In a study of treatment for asthma, Rutten-van Molken and co-workers (1994) suggest a log transformation of mean costs as well as adding a constant to the untransformed cost (in order to allow for the number of patients with costs of zero). Gray and colleagues (1997) transformed costs to natural logs before further analysis. Bouckaert and Crott (1997) suggest that the distribution of cost data may be bimodal, that is, with a peak at two points (e.g. zero and one other point). If parametric tests are to be used, the distribution of the data after transformation should be tested to assess whether the transformation has resulted in a distribution approximating normality (Coyle, 1996). A limitation of using transformations is that it is more difficult to interpret the transformed scale. An alternative way of deciding whether the use of a normal distribution and associated parametric tests is appropriate is to perform bootstrapping on the original data in order to test the robustness of the data to assumptions of normality. Bootstrapping, which allows distributions to be constructed empirically, has been used more often with cost-effectiveness ratios and is discussed below (pp. 30–31).

### *Uncertainty in resource use and cost*

Uncertainty in resource-use and cost differences is addressed by performing statistical tests of difference, while uncertainty in cost variation can be explored by using regression analysis. Both these approaches are discussed in this section.

As noted above, because a few patients may incur high-cost events, this results in cost data being skewed. Statistical tests may be used to compare costs between groups. The test used will depend on assumptions regarding the distribution of the data. Available parametric tests are: *t*-tests on untransformed data (which assume normality and that the variances of the two groups are equal); and *t*-tests on log transformed data (which require the data to be log normal and the variances in the log scale to be equal). Zhou and co-workers (1997) propose a method of comparing costs based on the standard normal distribution, using *Z*-scores to test the equality of mean costs between groups. For non-parametric data, Wilcoxon tests are used (on untransformed data). Non-parametric tests address differences in medians rather than means; therefore they may not be appropriate if the objective is to compare means and identify a mean difference in total cost (Coyle, 1996). Mean costs are usually preferred over median costs because they can be related back to total cost.

Debate is ongoing about whether statistical tests should be performed on the individual resource quantities or on total costs. This may not be appropriate because total costs are based on multiple parameters, reflecting the sum of many cost-generating events multiplied by their associated unit costs. Consequently, the variability in total cost is a function of the variability of each component, as well as the correlation among the components (Mullahy and Manning, 1995). Inconsistencies may therefore arise when testing for differences in single resource items. Coyle (1996) illustrates the case where components of cost are substitutes and hence are negatively correlated, and argues that testing for difference in single resource-use items is inappropriate. Some studies may also perform subgroup analysis (see for example, Mark *et al.*, 1995); that is, they test for differences in cost

between, for example, different age groups. Subgroup analysis is appropriate, however, only if there is sufficient statistical power to detect differences.

In studies testing for differences in resource use, non-parametric tests have been utilised, such as the Wilcoxon rank sum test (for continuous measures) or chi-squared tests (for discrete measures) (Gladman *et al.*, 1994; Menzin *et al.*, 1996; Sculpher *et al.*, 1996a). Others have used parametric tests based on untransformed data (Tarrier *et al.*, 1991) or parametric tests based on log transformed data (Burns *et al.*, 1993).

Confidence intervals should be presented around point estimates to indicate significant differences. The uncertainty in cost-generating events and total costs can be explored by using measures of variability. These measures, such as a standard deviation or variance, should be presented around the mean cost differences. Variability in cost-generating events can also be examined by calculating the variance/mean ratio (or overdispersion) of events (Spiegelhalter *et al.*, 1996). If the variance equals the mean, there is no overdispersion. Variability in costs can be examined by calculating the coefficient of variation, which expresses the standard deviation of a cost as a percentage of the mean (Gray *et al.*, 1997). This measure allows the variability between different types of cost to be compared.

Regression analysis has been used to examine the relationship between total costs and the factors determining total costs. These factors may include patients' characteristics at baseline. The advantage of using regression analysis is that it can control for these confounding effects (Dudley *et al.*, 1993; Oster *et al.*, 1995; Rutten-van Molken *et al.*, 1995; Schulman *et al.*, 1996a). In a multicentre study, centre effects may be an additional source of confounding factors (Coyle, 1996). Rutten-van Molken and co-workers (1994) performed a regression analysis to assess the true significance of the effect of treatment on costs, controlling for patient characteristics. They used a multiple regression model to examine the log transformed total costs between groups. The regression model was then used to predict the transformed costs for each group and standardise the difference in baseline characteristics. The predictions were retransformed from logged costs into costs. Differences in the costs at the baseline assessment can also be adjusted for by performing an analysis of covariance with the baseline costs as covariates (Gray *et al.*, 1997).

### *Distribution of the incremental cost-effectiveness ratio*

The incremental cost-effectiveness ratio is a ratio of two uncertain numbers, which introduces additional challenges into the statistical analysis of uncertainty in cost-effectiveness because the ratio is not normally distributed. The uncertainty surrounding the ratio may be larger than either element (Mullahy and Manning, 1995). Because the costs and effects are likely to be correlated, the incremental cost-effectiveness ratio is a ratio of two correlated random variables. The direction and magnitude of the covariance between costs and effects are empirical questions. It has been suggested that costs and effects may be negatively correlated because adverse effects are also likely to be more expensive, on average (Gold *et al.*, 1996). An additional complication arises where the difference in effects is zero or close to zero. In this case, the inverse of the ratio can cover a range from very small to quite large (Mullahy and Manning, 1995).

### *Uncertainty in the incremental cost-effectiveness ratio*

Traditionally, incremental cost-effectiveness ratios were presented as point estimates, without any indication of variability or uncertainty. More recently, the use of statistical analysis in examining the uncertainty in cost-effectiveness has been examined (O'Brien *et al.*, 1994; Mullahy and Manning, 1995; Siegel *et al.*, 1996; Laska *et al.*, 1997). Estimation of the uncertainty in the cost-effectiveness ratio requires estimating the variance in the ratio, which is a complex task, given the nature of the ratio. The variance of the estimator cannot be estimated exactly (O'Brien *et al.*, 1994; Rutten-van Molken *et al.*, 1994; Mullahy and Manning, 1995; Wakker and Klaassen, 1995; Briggs *et al.*, 1997).

Various methods have been suggested for calculating the confidence interval for the incremental cost-effectiveness ratio *(Box 13)*.

**BOX 13 Potential methods for calculating confidence intervals around cost-effectiveness ratios**

- simulation from summary statistics
- box method
- delta method
- method based on Fieller's theorem
- parametric or non-parametric bootstrapping
- calculate directly using logs.

First, simulation from summary statistics (estimates of the parameter and the variance–covariance matrix) can be used. Secondly, the box method can be performed, which uses separate confidence limits for costs and effects to produce a confidence interval for the cost-effectiveness ratio (Polsky *et al.*, 1997). A third procedure is the delta method, which uses a Taylor series approximation to estimate the variance of a cost-effectiveness ratio (O'Brien *et al.*, 1994). From this it is possible to derive a confidence interval around the cost-effectiveness ratio. Fourthly, a method based on Fieller's theorem can be used, which assumes that the numerator and denominator follow a normal distribution but does require the ratio to be distributed normally. Fifthly, parametric or non-parametric bootstrapping can be performed, which involves resampling from study data and computing cost-effectiveness ratios for each sample. The assumption of bootstrapping is that the distribution generated by the resampling process is a good estimate of the underlying distribution. Finally, the log of the cost-effectiveness ratio and its confidence interval can be calculated directly, but this is not possible if the differences in cost or effects are close to zero (Mullahy and Manning, 1995).

Several researchers have used at least one of these methods or several of them in empirical studies (Chaudhary and Stearns, 1996; Obenchain *et al.*, 1997; Polsky *et al.*, 1997). Chaudhary and Stearns (1996) compared several methods to obtain confidence intervals for incremental cost-effectiveness ratios and showed that a parametric bootstrap method was biased but, once it is adjusted for bias, it is computationally simpler than a non-parametric bootstrap. Stinnett (1996) showed that bootstrap simulation methods could be used to estimate bias in the cost-effectiveness ratio and to adjust accordingly. Polsky and co-workers (1997) found Fieller's and the non-parametric bootstrap methods to be more accurate. Others have suggested using Fieller's method (Gardiner *et al.*, 1995; Sacristan *et al.*, 1995; Sacristan 1996; Chaudhary and Stearns, 1996; Willan and O'Brien, 1996). Wakker and Klaassen (1995) used a method based on Bonferroni's inequality for obtaining one-sided confidence intervals. Chaudhary and Stearns (1996) argued that, before deciding on a method, the extent to which the data are consistent with the model assumptions should be considered.

Van Hout and colleagues (1994) calculated a 95% probability ellipse for costs and effects and constructed an acceptability curve for the cost-effectiveness ratio as an alternative estimate of precision. This meant that the probability of the ratio of costs and effects falling within the cost-acceptability curve would be 95%.

## Sensitivity analysis

The main method used to address uncertainty in economic evaluation is sensitivity analysis (Briggs *et al.*, 1994; Briggs and Sculpher 1995; Agro *et al.*, 1997). Even if statistical analysis has been performed, sensitivity analysis is still likely to be required in order to examine the extent of uncertainty in point estimates, such as the discount rate or unit cost data. It will still be required to examine uncertainty in results if data have been modelled or extrapolated.

Sensitivity analysis involves the systematic investigation of how changes in the selected assumptions affect the overall outcome and indicates which parameters in a model are critical to the results. A critical component of total cost or cost-effectiveness ratio is changed by some meaningful amount and the total cost or cost-effectiveness ratio is then recalculated. This gives some indication of how sensitive the results might be to a substantial, but not implausible, change in the values of a parameter of interest. Sensitivity analysis can be used: when no estimates of the parameter are available and informed guesses have to be made; where estimates are subject to debate because of known imprecision in the estimation procedure; or where there is methodological controversy (Drummond and Jefferson, 1996). The impact of changes in staff costs on total cost was examined by Hundley and co-workers (1995). The impact on cost-effectiveness of changes in a wide range of assumptions, including resource-use and unit cost estimates, was examined by Sculpher and colleagues (1996b).

### *Stages of sensitivity analysis*

There are several stages involved in sensitivity analysis. For most types, the parameters to be varied and the amount of variation (or ranges) to be imputed are selected. The specification of ranges across which to vary parameters is a key component and justification for the ranges imputed should be given (Briggs *et al.*, 1994). Ranges may be based on: clinically meaningful ranges; 95% confidence intervals for parameters; or highest and lowest ranges possible (Gold *et al.*, 1996). The amount of change in base results that is acceptable or constitutes a robust finding also has to be specified. All the stages of sensitivity analysis thus involve subjective elements.

### *Types of sensitivity analysis*

There are various forms of sensitivity analysis (Briggs *et al.*, 1994). These include univariate,

multivariate, extreme, threshold and probabilistic methods. Univariate or simple sensitivity analysis examines the impact that a change in a single parameter may have. It is useful in providing building blocks for further sensitivity analyses. Simple sensitivity may understate, or overstate, the overall uncertainty if parameters interact; it therefore provides an incomplete estimate of the uncertainty (Mullahy and Manning, 1995). In order to address interaction between parameters, multivariate sensitivity analysis can be performed, in which several parameters are changed simultaneously. This method is encouraged in the Canadian guidelines (Canadian Coordinating Office for Health Technology Assessment, 1994). Extreme sensitivity analysis identifies the combination of parameter values that yields the worst (highest) or best (lowest) cost-effectiveness. Extreme sensitivity analysis is useful only if results are insensitive to change, otherwise a large band of uncertainty surrounds the results (Gold *et al.*, 1996). Threshold analysis is a further form of sensitivity analysis, which determines the value a parameter would have to take in order for the most effective option to cease to dominate. It can be useful when high and low ranges of a parameter are difficult to specify. Probabilistic sensitivity analysis employs Monte Carlo analysis to simulate the model, with assumptions about variability in the data (Doubilet *et al.*, 1985). Repeated samples are drawn from these distributions to determine an empirical distribution (O'Brien *et al.*, 1994).

## Other analyses

### Generalising results

The generalisability of results from clinical trials is an issue that is important in terms of both clinical and economic data (Leidl, 1994). In clinical terms, the main concern of generalisability is the representativeness of the trial sample. In economic terms, factors limiting generalisability include differences in compliance, setting, organisational practice, capacity and economies of scale (Leidl, 1994; Baltussen *et al.*, 1996). Generalisability might also address various price levels, different settings and pooled data (Drummond and Davies, 1991). Data derived from the trial may therefore require adjustment in order to maximise external validity (Drummond *et al.*, 1997). Studies analysing this aspect have tended to focus on generalisability across countries (Drummond *et al.*, 1992; Jefferson *et al.*, 1996).

Most of the methods used to generalise results are based on sensitivity analysis (Baltussen *et al.*, 1996; Jacobs and Baladi, 1996; O'Brien, 1996). These analyses incorporate data from a variety of sources. If data have been collected at the design stage, then they can be used in the modelling process to consider the generalisability of results. If data are not available, a Bayesian approach to revising trial estimates could be adopted (Rittenhouse, 1997). Rittenhouse (1995) suggests incorporating data from a non-trial 'real world' environment and adjusting analysis so that the results are more likely to reflect clinical practice. Adjusting analyses from clinical trials to reflect effectiveness rather than efficacy may be problematic because there are no valid or standard means of making these adjustments.

In the context of multicentre or multinational clinical trials, the impact of any differences in resource use and unit cost across centres may be explored by using sensitivity analysis. For example, a single set of unit costs could be applied to centre- or country-specific resource-use data, or country-specific unit cost data could be applied to pooled resource-use data (Drummond and Davies, 1991; Drummond and Jefferson, 1996). This would assist in identifying the direction and magnitude of differences across centres and countries. In multinational studies, it may be possible to estimate country-specific cost-effectiveness ratios but, given the concerns about generalisability, a single cost-effectiveness ratio may not be relevant at local level.

### Modelling long-term outcomes

There is a growing interest in the use of models in economic evaluation (Luce, 1995; Office of Health Economics, 1997b; Rittenhouse, 1996b; Buxton *et al.*, 1997). The term 'model' is widely used to refer to decision analytical models, Markov models and various forms of regression analyses. In the context of clinical trials, modelling may be required to predict long-term outcomes beyond the end-point of the trial; this is the main application of modelling discussed here. Other forms of analysis may also be useful for handling missing and censored data, as discussed below (pp. 33–34).

The extrapolation of results from surrogate or intermediate outcomes to 'final' outcomes, or from short-term or long-term periods, is often required (Buxton *et al.*, 1997). Several types of model are available to accomplish this *(Box 14)*. The DEALE method (Declining Exponential Approximation to the Life Expectancy model) is the simplest technique for computing life expectancy and assumes that the hazard of dying from the disease and the hazard of dying normally are constant (Beck *et al.*, 1982). Markov

**BOX 14 Methods available for modelling long-term outcomes**

- DEALE method
- Markov models
- survival analysis
  - Cox proportional hazard modelling
  - Kaplan–Meier analysis
  - life table analysis.

models can also be used to analyse events that repeat and occur over time. The basic assumption of this model is that individuals are, at any time, in one of a finite set of states of health, and that health changes from state to state according to a set of transition probabilities (Beck and Pauker, 1983; Sonnenberg and Beck, 1993). Gompertz functions can be used to extrapolate survival from the baseline (Mark *et al.*, 1995). Survival analyses, such as Cox proportional hazard models (Mark *et al.*, 1995; Jonsson and Weinstein, 1997), Kaplan–Meier methods and life tables (Backhouse *et al.*, 1994; Hlatky *et al.*, 1997b) can also be used.

Schulman and co-workers (1991) used two different models to extrapolate clinical outcomes and survival beyond the trial for asymptomatic people with HIV infection who were treated with zidovudine, one based on a one-time benefit in the first year and the other on continuous benefit in future years. The sensitivity analysis showed that the cost-effectiveness ratio was extremely sensitive to the choice of model.

This highlights one of the limitations of modelling, the fact that assumptions made or the model used may affect the results to a large degree. Errors in modelling are important to identify because they may be cumulative; that is, any initial uncertainty surrounding an assumption may be multiplied through the model and lead to even greater uncertainty. If a model is to be used, it should be simple, presented in transparent form and validated if possible (Buxton *et al.*, 1997). Validation methods are discussed in an article by Sonnenberg and colleagues (1994).

## Analysis in the presence of missing and censored data

Further analysis may be required to pool the data and to handle missing and censored data. If data are from multicentre trials, they are likely to be heterogeneous. The methods used to pool data will depend on the design of the trial. If the trial is a cluster randomisation, where the unit of randomisation is, for example, the GP practice rather than the patient, heterogeneity could be taken into account by pooling differences using a random effects meta-analysis. This then produces larger standard errors than by assuming there are no differences between centres (Wonderling *et al.*, 1996b).

Possible approaches to use with missing data are non-random; they include ignoring missing data or attempting to estimate them by imputation. Regression, multiple regression and mean imputations are all possible forms of estimation (Little and Rubin, 1987). Replacement by mean imputation may be biased and reduce artificially the variance in parameter estimates (Coyle *et al.*, 1995; Little and Rubin, 1987). If data are missing non-randomly, then investigation of the missing cases may be required to examine whether there is any consistent pattern to them.

When censored data are present, approaches to handling this data also need to be found in order to estimate mean total cost. If the costs associated with censored observations are excluded, this may underestimate the costs (Rutten-van Molken *et al.*, 1995).

Several approaches have been used in empirical studies. One is to assign zero costs for the censoring period (Hlatky *et al.*, 1997a). An alternative is to carry forward the mean cost of all observations for patients with limited follow-up (Rutten-van Molken *et al.*, 1994, 1995). This has been argued to be appropriate because the high use of resources observed before withdrawal would have been sustained or may even have increased.

Statistical approaches used to handle censored clinical data have been adapted for use with economic data. Non-parametric methods, such as Kaplan–Meier survival analysis and life table analysis, have been suggested. Fenn and co-workers (1995) argue that, when calculating mean within-trial costs for samples of patients, Kaplan–Meier analysis should be used to adjust for censoring. In using this approach, costs, rather then survival time, are used as the dependent variable and are treated as right censored. It has been argued, however, that this approach is valid only if the additional cost per period is constant over time (Lin *et al.*, 1997). An alternative approach is to use the total cost for a specified period as the dependent variable. The period is chosen so as to reduce censored data. Lin and colleagues (1997) partitioned the survival curve into intervals and estimated costs for each interval, but found these estimates to be consistent only if censoring occurred solely at the boundaries of the interval.

There are several empirical applications of survival analysis to cost data, Kaplan–Meier analysis in particular, with the aim of taking account of censored data in order to estimate cumulative costs (Backhouse *et al.*, 1994; Etzioni *et al.*, 1996; Schulman *et al.*, 1996b; Gruger and Backhouse, 1997; Hlatky *et al.*, 1997a).

The underlying assumption of these approaches is one of independence between the variable of interest and its censoring mechanism, yet data may be non-randomly censored, for example, if patients who are more ill drop out of a trial. Loss to follow-up may therefore be related to treatment. Censoring may also be 'informative'; that is, the censoring time may provide information about future events (Glasziou *et al.*, 1990). If this is the case then these methods may not be appropriate. However, not adjusting for censored data may also lead to bias. The issues surrounding the suitable use of survival analysis have been reviewed in detail elsewhere in the context of quality of life data (Billingham *et al.*, 1999).

## Summary

Data analysis may also influence the design of the study. In summarising and synthesising cost data, issues such as how to pool data and how to handle missing and censored data have to be considered. It is important to address the variability in, and the distribution of, cost data. It is generally agreed that mean costs convey more useful information than medians because they relate to total cost. The procedures used to address uncertainty in methods and results include both statistical and sensitivity analyses; these have complementary roles. Sensitivity analysis can also be used to generalise results.

The main methodological issues when there is a lack of consensus include: the investigation of methods for handling missing and censored data; methods of generalising and/or adjusting results from clinical trials to routine practice; and exploring uncertainty in cost-effectiveness results.

# Chapter 5

# Presentation and reporting of results

## Common reporting format

A common reporting format for economic evaluations would both increase the transparency of methods and results, and facilitate comparison across studies (Drummond *et al.*, 1997). Reporting styles are influenced by the objectives of the study and the perspective of the analysis. Some reporting guidelines are national guidelines intended for the reimbursement of pharmaceuticals (Canadian Coordinating Office for Health Technology Assessment, 1994). Others are aimed at studies reporting their results in journals (Gold *et al.*, 1996; Drummond *et al.*, 1997). Despite differences in objectives, there is some common agreement on reporting styles (Drummond *et al.*, 1997). Detailed lists of items for inclusion in the reporting of economic evaluation are presented in the guidelines of the Canadian Coordinating Office for Health Technology Assessment (1994), by Gold and co-workers (1996), and by Drummond and colleagues (1997). For the purposes of this review, four broad headings (background information, methods, results and discussion) are used as the basis for discussing the presentation and reporting of results.

### Background information

Background information on the research objectives, description of comparators, study design, target population of the intervention, boundaries of analysis, time horizon and perspective adopted should all be reported (Gold *et al.*, 1996; Siegel *et al.*, 1996). Details of related study references, the target audience and descriptions of any pharmaceutical products should also be given (Canadian Coordinating Office for Health Technology Assessment, 1994). Guidelines for the reporting of clinical trials could also be consulted so that the reporting of medical evidence is consistent with current recommendations (Altman, 1996; Begg *et al.*, 1996). Patient characteristics may affect costs and effects; hence, patients' clinical and demographic characteristics should be reported (Rutten-van Molken *et al.*, 1995; Schulman *et al.*, 1996a; Hlatky *et al.*, 1997a). The details of where and when the study took place should also be stated. Baseline information on centre characteristics is also useful. Items of background information that should be reported are summarised in *Box 15*.

**BOX 15 Reporting of background information**

- research objectives
- description of comparators
- perspective
- time horizon of study
- patient characteristics
- centre characteristics.

### Methods

The methods used should be made transparent and the following information should be reported:

- descriptions of the costing methods used, including the types of cost and the reasons for adopting particular methods (Canadian Coordinating Office for Health Technology Assessment, 1994)
- sample size calculations and assumptions (Mason and Drummond, 1995)
- reasoning behind centre selection
- data collection instruments used and evidence on their validity and reliability
- timing and management of data collection
- when modelling is used: an explanation of the reported parameters modelled; the variables included or excluded in the modelling; the statistical relations assumed or derived; and whether the evidence supports the assumptions (Drummond and Jefferson, 1996)
- methods used to investigate uncertainty
- methods used to account for missing and censored data
- methods used to adjust for differential timing
- dates of the estimates of resource use and unit cost, together with any details of the adjustment to a more recent price level (Drummond *et al.*, 1993; Mason and Drummond, 1995)
- year and currency of cost data (Drummond and Jefferson, 1996).

### Results

Costs should be presented with some information on their variability. Means, rather than median costs, convey more useful information.

Disaggregation of results can aid transparency (Drummond *et al.*, 1993; Coyle *et al.*, 1995; Williams *et al.*, 1995) *(Box 16)*. Results can be

**BOX 16 Disaggregating results**

- costs disaggregated into resource-use and unit costs
- total costs disaggregated into components
- productivity costs disaggregated into resource quantities and valuation
- ratio disaggregated into total costs and total effects
- costs and effects disaggregated according to the groups affected.

disaggregated in a number of ways. First, costs should be disaggregated into components to indicate those costs contributing most to the total. Secondly, total costs should be separated into resource-use and unit cost estimates. Thirdly, productivity costs can be disaggregated into resource quantities and valuation by reporting days of work lost and the valuation of work lost separately (Drummond *et al.*, 1997). The major outcomes (e.g. the impact on quality of life) can be presented in a disaggregated as well as an aggregated form (Drummond and Jefferson, 1996). If a cost–consequence analysis has been conducted, then the total costs and consequences will be presented in disaggregated form anyway (for example, Nicholl *et al.*, 1992). Fourthly, even if the form of analysis is not cost–consequence, incremental costs and effects should be presented separately. This may provide more useful information to decision makers than an incremental cost-effectiveness ratio. Finally, the distributional implications of the results should be quantified by disaggregating the analysis of costs and effects according to the groups affected (Her Majesty's Treasury, 1997).

The incremental cost-effectiveness ratios can be presented graphically on a cost-effectiveness plane (Anderson *et al.*, 1986; Black, 1990) and the probability of an intervention being cost-effective can be shown using a cost-effectiveness acceptability curve (van Hout *et al.*, 1994).

When there are many cost items, reporting should concentrate on the main costs (Drummond and Jefferson, 1996). The separate reporting of fixed and variable costs is also recommended (Williams *et al.*, 1995). The composition of total cost can be shown graphically (Backhouse *et al.*, 1994; Spiegelhalter *et al.*, 1996). The costs should be presented undiscounted so that readers can apply their own discount rates (Drummond and Jefferson, 1996). Long- and short-term results should be given (Canadian Coordinating Office for Health Technology Assessment, 1994). Response rates and drop-out rates should also be reported (Mason and Drummond, 1995), and also the results from any modelling and sensitivity analysis.

### Discussion

The discussion should include comment on the limitations of the study (Gold *et al.*, 1996). The policy relevance and generalisability of results should also be discussed (Gold *et al.*, 1996; Drummond *et al.*, 1997). Any comparisons with evaluations of other health service interventions in other studies should be made only when a close similarity in study methods can be demonstrated (Drummond and Jefferson, 1996). There are often space constraints if the results of studies are published in journals; consequently, several authors recommend that technical reports should be made available to readers on request (Canadian Coordinating Office for Health Technology Assessment, 1994; Gold *et al.*, 1996; Drummond *et al.*, 1997).

## Summary

The presentation of results addresses the reporting formats used. Results should be presented in a disaggregated manner, for example, by separating resource use from unit costs and reporting the contribution of different types of cost to total cost. The development of a common reporting format for economic evaluations would increase the transparency of both methods and results. The design of future studies relies on transparent reporting in earlier studies so that issues such as the variability in cost data can be determined.

# Chapter 6
# Discussion

## Status of methodological issues

The recent increase in the number of economic evaluations being conducted alongside or as an integral part of clinical trials has provided the opportunity to collect and analyse patient-specific resource-use (and hence cost) data. These opportunities have in turn focused attention on a range of important methodological issues concerning the collection of resource-use data for costing purposes and its analysis. This systematic review has drawn on the existing methodological literature on conducting economic evaluation alongside clinical trials, with the aim of identifying the main methodological issues associated with costing data collection and analysis, and establishing whether these issues can be resolved by reference to the theory underlying economic evaluation or the practical experience of undertaking such studies.

As the collection of economic data within or alongside clinical trials is seen as a routine requirement, and decisions about data collection are made by investigators designing trials who may not have experience or formal training in health economics, the importance of clarifying, and if possible resolving, these issues increases.

Given that methodological approaches to the collection and analysis of resource-use data for costing are still developing, there is a danger in making recommendations on methods that may impose rigid standards and constrain further methodological development and debate. Some apparent disagreement about how to handle certain issues is inevitable. Different methods may be required to answer different economic questions, or the same question in different health service contexts. Furthermore, resolutions of many of the design decisions in costing, and in economic evaluation more generally, are fundamentally dependent on the theoretical perspective of the analyst and the perspective of the study. Such issues can only be resolved once there is a decision about the appropriate value system, viewpoint or theoretical standpoint for a particular study.

In discussing the status of current methodological issues, a useful way forward is to distinguish those where there is general agreement from those of disagreement. Issues where there is disagreement can be further divided into two types:

- those that reflect legitimate differences in values and perspective
- those, often more practical, issues that are amenable to further elucidation by empirical research.

It appears from this review that perhaps, not surprisingly, the literature and main methodological texts have focused more attention on the first type of issues of disagreement, while essentially ignoring the need to improve our understanding of, for example, the relative merits of alternative methods of data collection. The first type of issue of disagreement can be resolved only by agreement about what are the appropriate values and perspective in the context of a particular study or programme of research. The second could, and should, be resolved by formal empirical testing of alternative methods. The issues falling into these three categories (one general agreement and two general disagreement) have been identified throughout the review and are now summarised. Although the number of published and commissioned economic evaluations alongside clinical trials is increasing rapidly, and has done so since the literature for this systematic review was identified, some of these studies stress the importance of further research into methodological issues. Many of the issues identified in this review therefore remain relevant and further research into them would be timely.

### Methodological issues where there is general agreement

*Box 17* summarises the methodological issues where there is general agreement.

### Methodological issues remaining open because of legitimate differences in values and perspectives

*Box 18* summarises the main methodological issues that remain open because they depend

**BOX 17 Methodological issues where there is general agreement**

- identifying perspective of study
- measuring units of resource use and applying appropriate unit cost
- measurement of health service costs of the whole intervention
- need to use existing information or pretrial study to inform study design
- need to calculate sample sizes for costs/ cost-effectiveness
- analysis of uncertainty
- use of discounting
- importance of generalising results
- transparency in methods and results.

**BOX 18 Summary of major methodological issues remaining open because of legitimate differences in values and perspectives**

- which perspective to adopt
- whether to base choices on economic welfare theory
- which approach to analysis to adopt: estimation, hypothesis testing or decision analysis
- which productivity costs should be included
- which future costs should be included.

upon the perspective and values adopted in the study. All these have been widely debated in the literature. They could of course legitimately be determined by a user or funder of such studies prescribing what values that body wishes to see reflected in studies undertaken for it or with its funding. This indeed has been the case with the context-specific guidelines produced for example in Ontario (Canadian Coordinating Office for Health Technology Assessment, 1994) and Australia (Commonwealth of Australia, 1995) for economic evaluations required to support submissions to have new drugs reimbursed within those publicly-funded health care systems. As discussed in chapter 2 (pp. 8–9), there is disagreement about the scope of the measurement of productivity costs and future costs. Resolution of these issues requires more considered arguments about the basis upon which they should be included or excluded.

## Methodological issues requiring empirical testing

Many of the more detailed methodological issues relate to the best ways of designing data collection exercises, and relate to sources of information, levels of necessary detail, the frequency of collecting data, modes of data collection, and so on. In principle, for such issues it should be possible to make specific, or at least context-specific, recommendations based on past experience. Unfortunately, although there are successful studies that can be drawn on as examples of seemingly good practice, and certainly such examples should be considered in designing new trials, this review found few examples of experimental studies that tested the relative performance of different methods. Hence, it is probably premature to offer firm recommendations on methods for handling even these methodological issues. Empirical studies need to be undertaken formally to address these questions and those undertaking future applied studies should be encouraged to build into their design some specific testing of these methodological issues. The issues falling into this category, which should form part of a future research agenda, are summarised in *Box 19*.

**BOX 19 Summary of methodological issues requiring further empirical research**

**Design**

- measuring productivity costs
- implications of alternative criteria for attribution of costs
- exploring optimal sampling approaches
- issues surrounding multicentre clinical trials and handling centre effects
- sample sizes for cost-effectiveness.

**Data collection**

- determining appropriate recall periods for data collection
- determining the validity and reliability of resource-use data collection instruments
- development of standard questionnaires for patient costs.

**Data analysis**

- investigation of methods of handling missing and censored data
- identifying methods of generalising or adjusting results from clinical trials
- synthesising cost and effect end-points.

**Presentation**

- development of a common reporting format for economic evaluations.

## Recommendations

The recommendations that arise from this review are targeted at three main groups: investigators; funding bodies; and those responsible for ensuring high standards in reporting of studies. A list of issues for further empirical research is suggested.

### Recommendations for investigators

Recommendations for investigators are summarised in *Box 20*. In undertaking studies, investigators should consider, in a systematic way, the options for resource-use data collection and ensure that the rationale for the chosen design is made explicit and transparent. They should also ensure that, in determining the design, full use is made of existing evidence relating to costs of the technologies or treatments under study in order to aid the design of well-focused data collection. In doing so, investigators may be able to draw on data archives, if such are established (this is discussed further below). It is also recommended that investigators should test and assist in the further development of the structured framework for study design presented in appendix 3. If, when designing studies, they confront methodological alternatives that are untested or unresolved, they should consider whether these could be formally tested within the applied study. This would contribute to the reduction in uncertainty surrounding the relative merits of alternative approaches. In order to aid transparency of the results, investigators should aim to ensure that details of their data collection methods and instruments are readily available and make certain that they report the results in such a way that analysis is repeatable by other investigators. They should also be willing to share data from previous studies with other researchers who are designing costing studies in similar fields. They should be prepared to deposit work in data archives if such are set up and, in so doing, contribute to the critical appraisal of the usefulness of such archives.

### Recommendations for funding bodies

Recommendations for funding bodies are summarised in *Box 21*. Funding bodies should be prescriptive about the perspective to be adopted in the studies funded. This would ensure the maximal comparative value of the results. If they are unwilling to do so, they should at least require that investigators are explicit about the perspective adopted and justify their positions. A similar approach to that adopted by the panel on cost-effectiveness in health and medicine in the USA (Gold *et al.*, 1996) could be adopted. This involves specifying a reference case (they chose the societal perspective) but not precluding investigators from adopting other perspectives where appropriate.

In reviewing research proposals, funding bodies should ensure that an adequate process of decision making about economic data collection has taken place or is allowed for in the research proposal. This may involve recognising the usefulness (and cost) of the initial analysis of existing data or, alternatively, funding some specific data collection and/or pilot study. Funding bodies should assist in the process of encouraging investigators to use

**BOX 20 Summary of recommendations for investigators**

- ensure that the rationale for data collection methods is explicit and transparent
- ensure that full use is made of existing evidence relating to costs, drawing on data archives if such are established.
- test and assist in the further development of the structured framework (decision aid) for study design (presented in appendix 3)
- consider whether outstanding methodological issues could be tested formally within an applied study
- report results in such a way that the analysis is repeatable by other investigators
- be willing to share data
- deposit work in data archives if such are established
- contribute to the critical appraisal of the usefulness of such archives.

**BOX 21 Summary of recommendations for funding bodies**

- be prescriptive about the perspective to be adopted in studies funded (or require that investigators are explicit and justify their positions)
- ensure that an adequate process of decision making about resource-use data collection has taken place, or is allowed for, in research proposals
- recognise that this may require the analysis of existing data, some initial analysis, data collection or a pilot study
- assist in the process by the establishment of an archive of data collection instruments and of empirical data sets
- require researchers funded to deposit their work in these archives
- actively encourage the empirical testing of detailed methodological alternatives within applied studies
- commission research into alternative methods of unit cost estimation.

existing data sets to inform design by the establishment of archives of empirical data sets and data collection instruments, and requiring research funded by them to be deposited therein. They should also actively encourage the empirical testing of detailed methodological alternatives within applied studies. The review of unit cost estimation was not included in the terms of this review; funding bodies should therefore consider supporting research into alternative methods of the estimation of unit costs.

**BOX 22 Summary of recommendations for those responsible for ensuring high standards of reporting of studies**

- ensure that authors are explicit about the perspectives adopted and methods used
- recognise the importance of a clear explanation of the methods of resource-use data collection
- require authors to make available supplemental information on methods and instruments in technical reports.

## Recommendations for those responsible for ensuring high standards in the reporting of studies

These recommendations are summarised in *Box 22*. Journal editors, or others responsible for ensuring high standards in the reporting of studies, such as the Health Technology Assessment programme, should ensure that authors are explicit about the perspectives adopted and methods used. They should recognise the importance of a clear explanation of the methods employed in resource-use data collection. They should also recognise, however, that limits placed on the length of main reports (whether or not published in a journal) may often require authors to make available supplemental information on precise data collection methods and instruments in technical reports, both to referees and, subsequently, to readers.

## Recommendations for further empirical research

Recommendations for further empirical research were discussed earlier in this chapter *(Box 19)*. These issues should form part of a future research agenda and empirical studies need to be undertaken to address these questions formally.

# Acknowledgements

We would like to thank: Mike Drummond for advice throughout the project, for comments on earlier drafts and for arranging access to the data set used in appendix 2; Ann Raven for assistance in the early stages of the project; Claire Davey for help in developing the methods of the review; Mark Sculpher for comments on an earlier draft; and Dave Whynes for contributions made to appendix 3.

The review was developed in consultation with the experts listed in appendix 1. We thank them for taking the time to provide detailed and helpful comments, which have greatly improved this review. The specific authorship of appendices 2 and 3 is indicated on the contents page.

The views expressed, and any errors remaining, are, however, those of the authors.

# References


Aaronson LS, Burman ME, 1994. Focus on psychometrics. Use of health records in research: reliability and validity issues. *Res Nurs Health*;**17**:67–73.

Adams M, McCall N, Gray D, Orza M, Chalmers T, 1992. Economic analysis in randomized control trials. *Med Care*;**30**:231–43.

Agro KE, Bradley CA, Mittman N, Iskedijan M, 1997. Sensitivity analysis in health economic and pharmacoeconomic studies: an appraisal of the literature. *PharmacoEconomics*;**11**:75–88.

Altman D, 1996. Better reporting of randomised controlled trials: the CONSORT statement. *BMJ*;**313**:570–1.

Anderson JP, Bush JW, Chen M, Dolenc M, 1986. Policy space areas and properties of benefit–cost/utility analysis. *JAMA*;**255**:794–5.

Anie KA, Jones PW, Hilton SR, Anderson HR, 1996. A computer-assisted telephone interview technique for assessment of asthma morbidity and drug use in adult asthma. *J Clin Epidemiol*;**49**:653–6.

Association of the British Pharmaceutical Industry, 1994. Guidelines on good practice in the conduct of economic evaluation of medicines. London: ABPI.

Backhouse M, Mauskopf JA, Jones D, Wold DE, 1994. Economic outcomes of colfosceril palmitate rescue therapy in infants weighing 1250 g or more with respiratory distress syndrome. *PharmacoEconomics*;**6**:358–69.

Bailey KR, 1994. Generalizing the results of randomized clinical trials. *Controlled Clin Trials*;**15**:15–23.

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

Beck J, Pauker S, 1983. The Markov process in medical prognosis. *Med Decis Making*;**3**:419–58.

Beck J, Kassirer JP, Pauker S, 1982. A convenient approximation of life expectancy (the 'DEALE'). I: Validation of the method. *Am J Med*;**73**:883–8.

Beecham J, 1994. Collecting information: the client service receipt interview. *Ment Health Res Rev*;**1**:6–8.

Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, *et al.*, 1996. Improving the quality of reporting of randomized controlled trials. *JAMA*;**276**:637–9.

Belgian Society for PharmacoEpidemiology, 1995. A proposal for methodological guidelines for economic evaluation of pharmaceuticals. Gent: BSPE.

Bennett C, Armitage J, Buchner D, Gulati S, 1994a. Economic analysis in Phase III clinical cancer trials. *Cancer Invest*;**12**:336–42.

Bennett CL, Armitage JL, LeSage S, Gulati SC, Armitage JO, Gorin NC, 1994b. Economic analyses of clinical trials in cancer: are they helpful to policy makers? *Stem Cells*;**12**:424–9.

Billingham LJ, Abrams KR, Jones DR, 1999. Quality of life assessment and survival data. *Health Technol Assess* (in preparation).

Birch S, Donaldson C, 1987. Applications of cost benefit analysis to health care. *J Health Econ*;**6**:211–25.

Birch S, Gafni A, 1992. Cost effectiveness/utility analyses: do current decision rules lead us to where we want to be? *J Health Econ*;**11**:279–96.

Birch S, Gafni A, 1994. Cost-effectiveness ratios: in a league of their own. *Health Policy*;**28**:133–41.

Black WC, 1990. The cost-effectiveness plane: a graphic representation of cost-effectiveness. *Med Decis Making*;**10**:212–15.

Blair E, Burton S, 1987. Cognitive processes used by survey respondents to answer behavioural frequency questions. *J Consumer Res*;**14**:280–8.

Bloom BS, Hillman AL, Fendrick AM, Schwartz JS, 1993. A reappraisal of hepatitis B virus vaccination strategies using cost-effectiveness analysis. *Am J Physicians*;**118**:298–306.

Bonsel G, Rutten F, Uyl-de Groot C, 1993. Economic evaluation alongside cancer trials: methodological and practical aspects. *Eur J Cancer*;**29A**:S10–S14.

Bouckaert A, Crott R, 1997. The difference in mean costs as a pharmacoeconomic outcome variable: power considerations. *Controlled Clin Trials*;**18**:58–64.

Boyer GS, Templin DW, Goring WP, Cornoni-Huntley JC, Everett DF, Lawrence RC, *et al.,* 1995. Discrepancies between patient recall and the medical record. *Arch Intern Med*;**155**:1868–72.

Briggs A, Fenn P, 1997. Trying to better than average: a commentary on 'statistical inference of cost-effectiveness ratios'. *Health Econ*;**6**:491–5.

Briggs A, Sculpher M, 1995. Sensitivity analysis in economic evaluation: a review of published studies. *Health Econ*;**4**:355–71.

Briggs A, Sculpher M, Buxton M, 1994. Uncertainty in the economic evaluation of health care technologies: the role of sensitivity analysis. *Health Econ*;**3**:95–104.

Briggs AH, Wonderling DE, Mooney CZ, 1997. Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. *Health Econ*;**6**:327–40.

British Medical Association/Royal Pharmaceutical Society of Great Britain, March 1996. British national formulary. London: BMA/RPSGB.

Brook RH, Chassin MR, Fink A, Solomon DM, Kosecoff J, Park RE, 1986. A method for the detailed assessment of the appropriateness of medical technologies. *Int J Technol Assess Health Care*;**2**:53–64.

Brouwer WBF, Koopmanschap MA, Rutten FFH, 1997a. Productivity costs measurement through quality of life? A response to the recommendation of the Washington panel. *Health Econ*;**6**:253–9.

Brouwer WBF, Koopmanschap MA, Rutten FFH, 1997b. Productivity costs in cost-effectiveness analysis: numerator or denominator: a further discussion. *Health Econ*;**6**:511–14.

Brown JB, Adams ME, 1992. Patients as reliable reporters of medical care process: recall of ambulatory encounter events. *Med Care*;**30**:400–11.

Bryan S, Brown J, 1997. Extrapolation of cost-effectiveness information to local settings. Health Economics Research Group discussion paper, no. 17. Uxbridge: Brunel University, 1997. (Now published as: Bryan S, Brown J, 1998. Extrapolation of cost-effectiveness information to local settings. *J Health Serv Res Policy*;**3**:108–12.)

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

Bryan S, Buxton M, McKenna M, Ashton H, Scott A, 1995b. Private costs associated with abdominal aortic aneurysm screening: the importance of private travel and time costs. *J Med Screening*;**2**:62–6.

Bulpitt CJ, Fletcher AE, 1990a. Economic assessments in randomized controlled trials. *Med J Aust*;**153**:S16–S19.

Bulpitt CJ, Fletcher AE, 1990b. Measuring costs and financial benefits in randomized controlled trials. *Am Heart J*;**119**:766–71.

Burns T, Raftery J, Beadsmore A, McGuigan S, Dickson M, 1993. A controlled trial of home-based acute psychiatric services. II: Treatment patterns and costs. *Br J Psychiatry*;**163**:55–61.

Buxton MJ, Dubois DJ, Turner RR, Sculpher MJ, Robinson PA, Searcy C, 1991. Cost implications of alternative treatments for AIDS patients with cryptococcal meningitis. Comparison of fluconazole and amphotericin B-based therapies. *J Infect*;**23**:17–31.

Buxton MJ, Drummond MF, van Hout BA, Prince RL, Sheldon TA, Szucs T, *et al.*, 1997. Modelling in economic evaluation: an unavoidable fact of life. *Health Econ*;**6**:217–27.

Cairns J, 1992. Discounting and health benefits: another perspective. *Health Econ*;**1**:76–9.

Canadian Coordinating Office for Health Technology Assessment, 1994. Guidelines for economic evaluation of pharmaceuticals. Ottawa: CCOHTA.

Chaudhary MA, Stearns SC, 1996. Estimating confidence intervals for cost-effectiveness ratios: an example from a randomised trial. *Stat Med*;**15**:1447–58.

CHKS, 1996. Acute Care 1996: healthcare resource groups national statistics 1995/1996. Alcester: CHKS.

Clark RE, Teague GB, Ricketts SK, Bush PW, Keller AM, Zubkoff M, *et al.*, 1994. Measuring resource use in economic evaluations: determining the social costs of mental illness. *J Ment Health Admin*;**21**:32–41.

Claxton K, Posnett J, 1996. An economic approach to clinical trial design and research priority-setting. *Health Econ*;**5**:513–24.

Cleary PD, Jette AM, 1984. The validity of self-reported physician utilization measures. *Med Care*;**22**:796–803.

Cohen DJ, Breall JA, Ho KKL, Weintraub RM, 1993. Economics of elective coronary revascularization: comparison of costs and charges for conventional angioplasty, directional atherectomy, stenting and bypass surgery. *J Am Coll Cardiol*;**22**:1052–9.

Commonwealth of Australia, 1995. Guidelines for the pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee (including submissions involving economic analyses). Canberra: Australian Government Print Office.

Cook J, Richardson J, Street A, 1994. A cost utility analysis of treatment options for gallstone disease: methodological issues and results. *Health Econ*;**3**:157–68.

Copley-Merriman C, Lair TJ, 1994. Valuation of medical resource units collected in health economic studies. *Clin Ther*;**16**:553–68.

Copley-Merriman C, Egbuonu-Davis L, Kotsanos JG, Conforti P, Franson T, Gordon G, 1992. Clinical economics: a method for prospective health resource data collection. *PharmacoEconomics*;**1**:370–6.

Coyle D, 1996. Statistical analysis in pharmacoeconomic studies. *PharmacoEconomics*;**9**:506–16.

Coyle D, Drummond M, 1996. Analysing differences in the costs of treatment across centres within economic evaluations. Paper presented at the International Health Economists' Association Conference, Vancouver, May 1996.

Coyle D, Drummond MF, on behalf of the Medical Research Council CHART Steering Committee, 1997. Costs of conventional radical radiotherapy versus continuous hyperfractionated accelerated radiotherapy (CHART) in the treatment of patients with head and neck cancer or carcinoma of the bronchus. *Clin Oncol*;**9**:313–21.

Coyle D, Davies L, Drummond M, 1995. Trials and tribulations: conducting economic evaluations alongside clinical trials. Paper presented to the UK Health Economists' Study Group, January, Bristol 1995. (Now published as: Coyle D, Davies L, Drummond M, 1998. Trials and tribulations: emerging issues in designing economic evaluations alongside clinical trials. *Int J Technol Assess Health Care*;**14**:135–44.)

Crawford SL, Tennstedt SL, McKinlay JB, 1995. A comparison of analytic methods for nonrandom missingness of outcome data. *J Clin Epidemiol*;**48**:209–19.

Davidoff A, Powe NR, 1996. The role of perspective in defining measures for the economic evaluation of medical technology. *Int J Technol Assess Health Care*;**12**:9–21.

Davis CE, 1994. Generalizing from clinical trials. *Controlled Clin Trials*;**15**:11–14.

Detsky AS, Naglie G, 1990. A clinician's guide to cost-effectiveness analysis. *Ann Intern Med*;**113**:147–54.

Dex S, 1991. The reliability of recall data: a literature review. Working papers of the ESRC Research Centre on Micro-social Change – 1911. Colchester: University of Essex.

Donaldson C, 1990. The state of the art of costing health care for economic evaluation. *Commun Health Stud*;**14**:341–56.

Donaldson C, Shackley P, 1997. Economic evaluation. In: Detels R, Holland WW, McEwen J, Omenn GS, editors. Oxford textbook of public health. 3rd ed. Oxford: Oxford University Press, 850–71.

Donaldson C, Hundley V, McIntosh E, 1996. Using economics alongside clinical trials: why we cannot choose the evaluation technique in advance. *Health Econ* **5**:267–9.

Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ, 1985. Probabilistic sensitivity analysis using Monte Carlo simulation. *Med Decis Making*;**5**:157–77.

Dranove D, 1995. Measuring costs. In: Sloan FA, editor. Valuing health care: cost, benefits and effectiveness of pharmaceuticals and other medical technologies. Cambridge: Cambridge University Press, pp. 61–75.

Drummond M, 1994. Economic analysis alongside controlled trials. Leeds: Department of Health.

Drummond M, 1995. Economic analysis alongside clinical trials: problems and potential. *J Rheumatol*;**22**:1403–7.

Drummond M, 1996. Experimental versus observational data in the economic evaluation of health technologies. Paper presented at the second CEPOR conference, University of North Carolina, April 1996. (Now published as: Drummond M, 1998. Experimental versus observational data in the economic evaluation of pharmaceuticals. *Med Decis Making*;**18**:S12–S18.)

Drummond MF, Coyle D, 1997. The role of pilot studies in the economic evaluation of health technologies. University of York, Mimeo 1997. (Now published as: Drummond MF, Coyle D, 1998. The role of pilot studies in the economic evaluation of health technologies. *Int J Technol Assess Health Care*;**14**:405–18.)

Drummond MF, Davies L, 1991. Economic analysis alongside clinical trials: revisiting the methodological issues. *Int J Technol Assess Health Care*;**7**:561–73.

Drummond MF, Jefferson TO, on behalf of the BMJ Economic Evaluation Working Party, 1996. Guidelines for authors and peer reviewers of economic submissions to the *BMJ*. *BMJ*;**313**:275–83.

Drummond M, O'Brien B, 1993. Clinical importance, statistical significance and the assessment of economic and quality-of-life outcomes. *Health Econ*;**2**:205–12.

Drummond MF, Stoddart GL, 1984. Economic analysis and clinical trials. *Controlled Clin Trials*;**5**:115–28.

Drummond MF, Stoddart GL, Torrance GW, 1987. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press.

Drummond MF, Mohide EA, Tew M, Streiner DL, Pringle DM, Gilbert JR, 1991. Economic evaluation of a support program for caregivers of demented elderly. *Int J Technol Assess Health Care*;**7**:209–19.

Drummond MF, Bloom BS, Carrin G, Hillman AL, Hutchings HC, Knill-Jones RP, *et al.*, 1992. Issues in the cross national assessment of health technology. *Int J Technol Assess Health Care*;**8**:671–82.

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

Drummond MF, O'Brien B, Stoddart GL, Torrance GW, 1997. Methods for the economic evaluation of health care programmes. 2nd ed. Oxford: Oxford University Press.

Dubois DJ, Buxton MJ, Turner RR, Leeming MRG, 1993. Incorporating quality of life and economic outcomes into multinational trials. Paris: Masson.

Dudley RA, Harrell FE, Smith LR, Mark DB, Califf RM, 1993. Comparison of analytic models for estimating the effect of clinical factors on the cost of coronary artery bypass graft surgery. *J Clin Epidemiol*;**46**:261–71.

Ellwein LB, Drummond MF, 1996. Economic analysis alongside clinical trials: bias in the assessment of economic outcomes. *Int J Technol Assess Health Care*;**12**:691–7.

Etzioni R, Urban N, Baker M, 1996. Estimating the costs attributable to a disease with application to ovarian cancer. *J Clin Epidemiol*;**49**:95–103.

Evans C, 1997. The use of consensus methods and expert panels in pharmacoeconomic studies. Practical applications and methodological shortcomings. *PharmacoEconomics*;**12**:121–9.

Evans DB, 1990. Principles involved in costing. *Med J Aust*;**153**(Suppl):S10–S12.

Fayers P, 1995. MRC quality of life studies using a daily diary card – practical lessons learned from cancer trials. *Quality Life Res*;**4**:343–52.

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

Fenn P, McGuire A, Phillips V, Backhouse M, Jones D, 1995. The analysis of censored treatment cost data in economic evaluation. *Med Care*;**33**:851–63.

Fenn P, McGuire A, Backhouse M, Jones D, 1996. Modelling programme costs in economic evaluation. *J Health Econ*;**15**:115–25.

Flannelly G, Campbell MK, Meldrum P, Torgerson DJ, Templeton A, Kitchener HC, 1997. Immediate colposcopy or cytological surveillance for women with mild dyskaryosis: a cost effectiveness analysis. *J Public Health Med*;**19**:419–23.

Gafni A, Birch S, 1993. Guidelines for the adoption of new technologies: a prescription of uncontrolled growth in expenditures and how to avoid the problem. *Can Med Assoc J*;**148**:913–84.

Garber AM, Phelps CE, 1997. Economic foundations of cost-effectiveness analysis. *J Health Econ*;**16**:1–31.

Gardiner J, Hogan A, Holmes-Rovner M, Rovner D, Griffith L, Kupersmith J, 1995. Confidence intervals for cost-effectiveness ratios. *Med Decis Making*;**15**:254–63.

Gardner M, Altman DG, 1986. Confidence intervals rather than *P* values: estimation rather than hypothesis testing. *BMJ*;**292**:746–50.

Gerard K, Mooney G, 1993. QALY league tables: handle with care. *Health Econ*;**2**:59–64.

Gladman J, Whynes D, Lincoln N, for the Domino Study Group, 1994. Cost comparison of domiciliary and hospital based stroke rehabilitation. *Age Ageing*;**23**:241–5.

Glasziou PP, Simes RJ, Gelber RD, 1990. Quality adjusted survival analysis. *Stat Med*;**9**:1259–76.

Glasziou PP, Simes RJ, Hall J, Donaldson C, 1997. Design of a cost-effectiveness study within a randomized trial: the LIPID trial for secondary prevention of IHD. *Controlled Clin Trials*;**18**:464–76.

Glick HA, Eisenberg JM, Koffer H, Kelley MA, Connell ML, 1988. Are cost-effectiveness analyses performed in clinical trials applicable to usual care?: a case study. *Med Decis Making*;**8**:328.

Goddard M, Hutton J, 1991. Economic evaluation of trends in cancer therapy: marginal or average costs. *Int J Technol Assess Health Care*;**7**:594–603.

Gold MR, Siegel JE, Russell LB, Weinstein MC, 1996. Cost-effectiveness in health and medicine. New York: Oxford University Press.

Gorsky RD, 1996. A method to measure the costs of counseling for HIV prevention. *Public Health Rep*;**111**:115–22.

Gray AM, Marshall M, Lockwood A, Morris J, 1997. Problems in conducting economic evaluations alongside clinical trials: lessons from a study of case management for people with mental disorders. *Br J Psychiatry*;**170**:47–52.

Gruger J, Backhouse M, 1997. Economic evaluation of antiviral therapy for the treatment of herpes zoster in immunocompetent adults. *PharmacoEconomics*;**11**:262–73.

Harlow SD, Linet MS, 1989. Agreement between questionnaire data and medical records. The evidence for accuracy of recall. *Am J Epidemiol*;**129**:233–48.

Hebert R, Bravo G, Korner-Bitensky N, Voyer L, 1996. Refusal and information bias associated with postal questionnaires and face-to-face interviews in very elderly subjects. *J Clin Epidemiol*;**49**:373–81.

Her Majesty's Treasury, 1997. Appraisal and evaluation in central government: treasury guidance. London: The Stationery Office.

Hlatky MA, Boothroyd DB, Johnstone IM, Marcus FI, Hahn E, Hartz V, *et al.*, 1997a. Long-term cost-effectiveness of alternative management strategies for patients with life-threatening ventricular arrhythmias. *J Clin Epidemiol*;**50**:185–93.

Hlatky MA, Rogers WJ, Johnstone I, Boothroyd D, Brooks MM, Pitt B, 1997b. Medical care costs and quality of life after randomization to coronary angioplasty or coronary bypass surgery. *N Engl J Med*;**336**:92–9.

Howard S, Normand C, Mugford M, Elbourne D, Field D, Johnson A, *et al.*, 1995. Costing neonatal care alongside the collaborative ECMO trial: how much primary research is required? *Health Econ*;**4**:265–71.

Howard S, Mugford M, Normand C, Elbourne D, Grant A, Field D, *et al.*, 1996. A cost-effectiveness analysis of neonatal ECMO using existing evidence. *Int J Technol Assess Health Care*;**12**:80–92.

Hughes J, Ryan M, Hinshaw K, Henshaw R, Rispin R, Templeton A, 1996. The costs of treating miscarriage: a comparison of medical and surgical management. *Br J Obstet Gynaecol*;**103**:1217–21.

Hundley VA, Donaldson C, Lang GD, Cruickshank FM, Glazener CMA, Milne JM, *et al.*, 1995. Costs of intra-partum care in a midwife-managed delivery unit and a consultant-led labour ward. *Midwifery*;**11**:103–9.

Hurley SF, Livingston PM, Thane N, Quang L, 1994. Mammographic screening: measurement of the cost in a population based programme in Victoria, Australia. *J Epidemiol Community Health*;**48**:391–9.

Hurley SF, Bond LM, Carlin JB, Evans DB, Kaldor JM, 1995. A method for estimating baseline health care costs. *J Epidemiol Community Health*;**49**:525–31.

Jacobs J, Bachynsky J, 1996. Costing methods in the Canadian literature on the economic evaluation of health care: a survey and assessment. *Int J Technol Assess Health Care*;**12**:721–34.

Jacobs P, Baladi JF, 1996. Biases in cost measurement for economic evaluation studies in health care. *Health Econ*;**5**:525–9.

Jefferson T, Mugford M, Gray A, Demicheli V, 1996. An exercise on the feasibility of carrying out secondary economic analyses. *Health Econ*;**5**:155–65.

Jobe JB, White AA, Kelley CL, Mingay DJ, Sanchez MJ, Loftus EF, 1990. Recall strategies and memory for health-care visits. *Milbank Q*;**68**:171–89.

Johannesson M, 1994. The concept of cost in the economic evaluation of health care. *Int J Technol Assess Health Care*;**10**:675–82.

Johannesson M, 1995. A note on the depreciation of the societal perspective in economic evaluation of health care. *Health Policy*;**33**:59–66.

Johannesson M, 1997. Avoiding double-counting in pharmacoeconomics studies. *PharmacoEconomics*;**11**:385–8.

Johannesson M, Karlsson G, 1997. The friction cost method: a comment. *J Health Econ*;**16**:249–55.

Johannesson M, Meltzer D, O'Conor R, 1997. Incorporating future costs in medical cost-effectiveness analysis. Implications for the cost-effectiveness of the treatment of hypertension. *Med Decis Making*;**17**:382–9.

Johnston K, Gerard K, Brown J, 1997. Generalising costs from the breast screening frequency and age trials. (Health Economics Research Group Discussion Paper no. 19.) Uxbridge: Brunel University.

Jones DA, 1995. The role of probability distributions in economic evaluations. *Br J Med Econ*;**8**:137–46.

Jonsson B, Bebbington PE, 1994. What price depression? The cost of depression and the cost-effectiveness of pharmacological treatment. *Br J Psychiatry*;**164**:665–73.

Jonsson B, Weinstein MC, 1997. Economic evaluation alongside multinational clinical trials: study considerations for GUSTO 11b. *Int J Technol Assess Health Care*;**13**:49–58.

Karlsson G, Johanesson M, 1996. The decision rules of cost-effectiveness analysis. *PharmacoEconomics*;**9**:113–20.

Kesteloot K, Penninckx F, 1993. The costs and effects of open versus laparoscopic cholecystectomies. *Health Econ*;**2**:303–12.

Knapp M, Beecham J, 1990. Costing mental health services. *Psychol Med*;**20**:893–908.

Knapp M, Beecham J, 1993. Reduced list costings: examination of an informed short cut in mental health research. *Health Econ*;**2**:313–22.

Knapp M, Beecham J, Koutsogeorgopoulu V, Hallam A, Fenyo A, Marks IM, *et al.,* 1994. Service use and costs of home based versus hospital based care for people with serious mental illness. *Br J Psychiatry*;**165**:195–203.

Koopmanschap MA, Rutten FFH, 1993. Indirect costs in economic studies: confronting the confusion. *PharmacoEconomics*;**4**:446–54.

Koopmanschap MA, Rutten FFH, 1996. A practical guide for calculating indirect costs of disease. *PharmacoEconomics*;**10**:460–6.

Koopmanschap MA, Rutten FFH, van Inveld BM, van Roijen L, 1995. The friction cost method for measuring indirect costs of disease. *J Health Econ*;**14**:171–89.

Koopmanschap M, Rutten FFH, van Inveld BM, van Roijen L, 1997. Reply to Johannesson's and Karlsson's comment. *J Health Econ*;**16**:257–9.

Krahn M, Gafni A, 1993. Discounting in the economic evaluation of health care interventions. *Med Care*;**31**:403–18.

Kuykendall DH, Johnson ML, 1995. Administrative databases, case-mix adjustments and hospital resource use: the appropriateness of controlling patient characteristics. *J Clin Epidemiol*;**48**:423–30.

Langham S, Thorogood M, Normand C, Muir J, Jones L, Fowler G, 1996. Costs and cost effectiveness of health checks conducted by nurses in primary care: the Oxcheck study. *BMJ*;**312**:1265–8.

Langkilde LK, Sogaard J, 1997. The adjustment of cost measurement to account for learning. *Health Econ*;**6**:83–5.

Laska EM, Meisner M, Siegel C, 1997. Statistical inference for cost-effectiveness ratios. *Health Econ*;**6**:229–42.

Laupacis A, Feeny D, Detsky AS, Tugwell PX, 1992. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. *Can Med Assoc J*;**146**:473–81.

Lave JR, Pashos CL, Anderson GF, Brailer D, Bubolz T, Conrad D, *et al.,* 1994. Costing medical care: using Medicare administrative data. *Med Care*;**32**:JS77–JS89.

Leidl RM, 1994. Some factors to consider when using the results of economic evaluation studies at the population level. *Int J Technol Assess Health Care*;**10**:467–78.

Leigh Brown AP, Lawrie ER, Kennedy ADM, Webb AG, 1997. Cost effectiveness of a prize draw on response to a postal questionnaire: results of a randomised trial among orthopaedic outpatients in Edinburgh. *J Epidemiol Community Health*;**51**:463–4.

Lin DY, Feuer EJ, Etzioni R, Wax Y, 1997. Estimating medical costs from incomplete follow-up data. *Biometrics*;**53**:419–34.

Linet MS, Harlow SD, McLaughlin JK, McCaffrey LD, 1989. A comparison of interview data and medical records for previous medical conditions and surgery. *J Clin Epidemiol*;**42**:1207–13.

Little R, Rubin D, 1987. Statistical analysis with missing data. New York: Wiley.

Luce BR, 1995. Policy implications of modeling the cost-effectiveness of health care technologies. *Drug Information J*;**29**:1469–75.

Luce BR, Elixhauser A, 1990a. Estimating costs in the economic evaluation of medical technologies. *Int J Technol Assess Health Care*;**6**:57–75.

Luce BR, Elixhauser A, 1990b. Standards for socioeconomic evaluation of health care products and services. Berlin: Springer-Verlag.

Mark DB, Naylor D, Hlatky MA, Califf RM, Topol EJ, 1994. Use of medical resources and quality of life after acute myocardial infarction in Canada and the United States. *N Engl J Med*;**331**:1130–5.

Mark DB, Hlatky MA, Califf RM, 1995. Cost-effectiveness of thrombolytic therapy with tissue plasminogen activator as compared with strepokinase for acute myocardial infarction. *N Engl J Med*;**332**:1418–24.

Mason J, 1997. The generalisability of pharmacoeconomic studies. *PharmacoEconomics*;**11**:503–14.

Mason J, Drummond M, 1995. Reporting guidelines for economic studies. *Health Econ*;**4**:85–94.

Mauskopf J, Schulman K, Bell L, Glick H, 1996. A strategy for collecting pharmacoeconomic data during Phase II/III clinical trials. *PharmacoEconomics*;**9**:264–77.

Meltzer D, 1997. Accounting for future costs in medical cost-effectiveness analysis. *J Health Econ*;**16**:33–64.

Menzin J, Oster G, Davies L, Drummond MF, Greiner W, Lucioni C, *et al.,* 1996. A multinational economic evaluation of rhDNase in the treatment of cystic fibrosis. *Int J Technol Assess Health Care*;**12**:52–61.

Ministry of Health, 1994. Ontario guidelines for economic analysis of pharmaceutical products. Ontario: Ministry of Health.

Morris J, Goddard M, 1992. Economic evaluation and quality of life assessments in cancer clinical trials: the CHART trial. *Eur J Cancer*;**29A**:766–70.

Morris J, Goddard M, Coyle D, Drummond M, 1993. Measuring cost and quality of life in radiotherapy trial treatments. (Centre for Health Economics, discussion paper 112.) York: University of York.

Morris S, McGuire A, Caro J, Pettitt D, 1997. Strategies for the management of hypercholesterolaemia: a systematic review of the cost-effectiveness literature. *J Health Serv Res Policy*;**2**:231–50.

Mugford M, 1995. The cost of neonatal care: reviewing the evidence. *Soc Prev Med*;**40**:361–8.

Mullahy J, 1996. What you don't know can't hurt you? Statistical issues and standards for medical technology evaluation. *Med Care*;**34**:DS124–35.

Mullahy J, Manning W, 1995. Statistical issues in cost-effectiveness analysis. In: Sloan FA, editor. Valuing health care: costs benefits and effectiveness of pharmaceutical and other medical technologies. Cambridge: Cambridge University Press, pp. 149–84.

Mushlin AI, Fintor L, 1992. Is screening for breast cancer cost-effective? *Cancer*;**69**(7 Suppl):1957–62.

Netten A, Dennett J, 1996. Unit costs of health and social care. Canterbury: Personal Social Services Research Unit, University of Kent.

Neumann P, Johannesson M, 1994. From principle to policy: using cost-effectiveness analysis. *Health Affairs*;**13**:206–14.

Nicholl JP, Brazier JE, Milner PC, Westlake L, Kohler B, Williams BT, *et al.,* 1992. Randomised controlled trial of cost-effectiveness of lithotripsy and open cholecystectomy as treatments for gallbladder stones. *Lancet*;**340**:801–7.

Obenchain RL, Melfi CA, Croghan TW, Buesching DP, 1997. Bootstrap analyses of cost effectiveness in antidepressant pharmacotherapy. *PharmacoEconomics*;**11**:464–72.

O'Brien B, 1996. Economic evaluation of pharmaceuticals: Frankenstein's monster or vampire of trials? *Med Care*;**34**:DS99–DS108.

O'Brien BJ, Drummond MF, Labelle RJ, Willan A, 1994. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Med Care*;**32**:150–63.

O'Brien B, Goeree R, Mohamed AH, Hunt R, 1995. Cost-effectiveness of *helicobacter pylori* eradication for the long-term management of duodenal ulcer in Canada. *Arch Intern Med*;**155**:1958–64.

Office of Health Economics, 1997a. Compendium of statistics. 11th ed. London: OHE.

Office of Health Economics, 1997b. The pros and cons of modelling in economic evaluation. *OHE Briefing*;**33**:1–12.

Olsen JA, 1994. Production gains: should they count in health care evaluations? *Scott J Polit Econ*;**41**:69–84.

Oster G, Huse DM, Lacey MJ, Regan MM, Fuchs HJ, 1995. Effects of recombinant human DNase therapy on health care use and costs in patients with cystic fibrosis. *Ann Pharmacother*;**29**:459–64.

Paganini-Hill A, Ross RK, 1982. Reliability of recall of drug usage and other health-related information. *Am J Epidemiol*;**116**:114–22.

Parsonage M, Neuberger H, 1992. Discounting and health benefits. *Health Econ*;**1**:71–6.

Phelps C, 1997. Good technologies gone bad: how and why the cost-effectiveness of a medical intervention changes for different populations. *Med Decis Making*;**17**:107–17.

Phillips K, Lowe E, Kahn J, 1994. The cost-effectiveness of HIV testing of physicians and dentists in the United States. *JAMA*;**271**:851–8.

Polsky D, Glick HA, Willke R, Schulman K, 1997. Confidence intervals for cost-effectiveness ratios: a comparison of four methods. *Health Econ*;**6**:243–52.

Posnett J, Jan S, 1996. Indirect cost in economic evaluation: the opportunity cost of unpaid inputs. *Health Econ*;**5**:13–23.

Powe N, Griffiths R, 1995. The clinical-economic trial: promise, problems and challenges. *Controlled Clin Trials*;**16**:377–94.

Ratcliffe J, 1995. The measurement of indirect costs and benefits in health care evaluation: a critical review. *Project Appraisal*;**10**:13–18.

Ratcliffe J, Ryan M, Tucker J, 1996. The costs of alternative types of routine antenatal care for low-risk women: shared care vs care by general practitioners and community midwives. *J Health Serv Res Policy*;**1**:135–40.

Revicki DA, Irwin D, Reblando J, Simon GE, 1994. The accuracy of self-reported disability days. *Med Care*;**32**:401–4.

Rigby K, Silagy C, Crockett A, 1996. Can resource use be extracted from randomized controlled trials to calculate costs? A review of smoking cessation interventions in general practice. *Int J Technol Assess Health Care*;**12**:714–20.

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

Rittenhouse BE, 1996a. Respondent-specific information from the randomised response interview: compliance assessment. *J Clin Epidemiol*;**49**:545–9.

Rittenhouse B, 1996b. Uses of models in economic evaluations of medicines and other health technologies. London: Office of Health Economics.

Rittenhouse B, 1996c. A novel compliance assessment technique: the randomized response interview. *Int J Technol Assess Health Care*;**12**:498–510.

Rittenhouse B, 1997. Exorcising protocol-induced spirits: making the clinical trial relevant for economics. *Med Decis Making*;**17**:331–9.

Rubins HB, 1994. From clinical trials to clinical practice: generalizing from participant to patient. *Controlled Clin Trials*;**15**:7–10.

Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC, for the Panel on Cost-effectiveness in Health and Medicine, 1996. The role of cost-effectiveness analysis in health and medicine. *JAMA*;**276**:1172–7.

Rutten-van Molken MPMH, Van Doorslaer EKA, Van Vliet RCJA, 1994. Statistical analysis of cost outcomes in a randomised controlled clinical trial. *Health Econ*;**3**:333–45.

Rutten-van Molken MPMH, Van Doorslaer EKA, Jansen MCC, Kerstijens HAM, Rutten FFH, 1995. Costs and effects of inhaled corticosteroids and bronchodilators in asthma and chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*;**151**:975–82.

Sacristan JA, 1996. Use of confidence intervals in health economic studies. *PharmacoEconomics*;**10**:432–3.

Sacristan JA, Day SJ, Navarro O, Ramos J, Hernandez JM, 1995. Use of confidence intervals and sample sizes in health economic studies. *Ann Pharmacother*;**29**:719–25.

Safran C, 1991. Using routinely collected data for clinical research. *Stat Med*;**10**:559–64.

Schulman KA, Lynn L, Glick HA, Eisenberg JM, 1991. Cost effectiveness of low-dose zidovudine therapy for asymptomatic patients with human immunodeficiency (HIV) infection. *Ann Intern Med*;**114**:798–802.

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

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

Sculpher MJ, Buxton MJ, 1993. The private costs incurred when patients visit screening clinics: the cases of screening for breast cancer and for diabetic retinopathy. (Health Economics Research Group discussion paper.) Uxbridge: Brunel University.

Sculpher MJ, Buxton MJ, Ferguson BA, Spiegelhalter DJ, Kirby AJ, 1992. Screening for diabetic retinopathy: a relative cost-effectiveness analysis of alternative modalities and strategies. *Health Econ*;**1**:39–51.

Sculpher MJ, Seed P, Henderson RA, Buxton MJ, Pocock S, Parker J, 1994. Health service costs of coronary angioplasty and coronary artery bypass surgery: the Randomised Intervention Treatment of Angina (RITA) trial. *Lancet*;**344**:927–30.

Sculpher MJ, Dwyer N, Byford S, Stirrat G, 1996a. Randomised trial comparing hysterectomy and transcervical endometrial resection: effect on health related quality of life and costs two years after surgery. *Br J Obstet Gynaecol*;**103**:142–9.

Sculpher MJ, Michaels J, McKenna M, Minor J, 1996b. A cost–utility analysis of laser assisted angioplasty for peripheral arterial occlusions. *Int J Technol Assess Health Care*;**12**:104–25.

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

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

Siegel C, Laska E, Meisner M, 1996. Statistical methods for cost-effectiveness analyses. *Controlled Clin Trials*;**17**:387–406.

Siegel JE, Weinstein MC, Russell LB, Gold MR, for the Panel on Cost-Effectiveness in Health and Medicine, 1996. Recommendations for reporting cost-effectiveness analyses. *JAMA*;**276**:1339–41.

Simon G, Wagner E, Vonkorff M, 1995. Cost-effectiveness comparisons using 'real world' randomised trials: the case of new antidepressant drugs. *J Clin Epidemiol*;**48**:363–73.

Simpson CA, Souney PF, 1988. Calculating costs of drug therapy: a review of a decade of literature (1977–1987). *Clin Ther*;**10**:616–50.

Smith K, Wright K, 1994. Informal care and economic appraisal: a discussion of possible methodological approaches. *Health Econ*;**3**:137–48.

Sonnenberg FA, Beck JR, 1993. Markov models in medical decision making. *Med Decis Making*;**13**:322–38.

Sonnenberg FA, Roberts MS, Tsevat J, Wong JB, Barry M, Kent DL, 1994. Toward a peer review process for medical decision analysis models. *Med Care*;**32**:JS52–JS64.

Spiegelhalter DJ, Jones DR, Parmar MKB, Gore SM, Fitzpatrick R, Cox DR, 1996. Being economical with the costs. (Department of Epidemiology and Public Health Technical Paper 96-06.) Leicester: University of Leicester.

Stinnett AA, 1996. Adjusting for bias in C/E ratio estimates. *Health Econ*;**5**:470–2.

Stone DH, 1993. Design a questionnaire. *BMJ*;**307**:1264–6.

Streiner DL, Norman GR, 1989. Health measurement scales: a practical guide to their development and use. Oxford: Oxford University Press.

Sudman S, Bradburn NM, 1982. Asking questions. London: Jossey Bass.

Tarrier N, Lowson K, Barowclough C, 1991. Some aspects of family interventions in schizophrenia. II: Financial considerations. *Br J Psychiatry*;**159**:481–4.

Torgerson DJ, Donaldson C, Reid DM, 1994. Private versus social opportunity cost of time: valuing time in the demand for health care. *Health Econ*;**3**:149–55.

Torgerson DJ, Ryan M, Ratcliffe J, 1995. Economics in sample size determination for clinical trials. *Q J Med*;**88**:517–21.

van Bergen PFMM, Jonker JJC, van Hout BA, van Domburg RT, Deckers JW, Azar AJ, *et al.,* 1995. Costs and effects of long-term oral anticoagulant treatment after myocardial infarction. *JAMA*;**273**:925–8.

van Hout BA, Al MJ, Gordon GS, Rutten FFH, 1994. Costs, effects and C/E-ratios alongside a clinical trial. *Health Econ*;**3**:309–19.

van Roijen L, Essink-Bot ML, Koopmanschap MA, Bonsel G, Rutten FFH, 1996. Labor and health status in economic evaluation of health care: the health and labor questionnaire. *Int J Technol Assess Health Care*;**12**:405–15.

Verbrugge LM, 1980. Health diaries. *Med Care*;**XVIII**:73–95.

Wakker P, Klaassen MP, 1995. Confidence intervals for cost/effectiveness ratios. *Health Econ*;**4**:373–81.

Weinberger M, Gold DT, Divine GW, Cowper PA, Hodgson LG, Schreiner PJ, *et al.,* 1993. Expenditures in caring for patients with dementia who live at home. *Am J Public Health*;**83**:338–41.

Weinstein MC, 1990. Principles of cost-effective resource allocation in health care organisations. *Int J Technol Assess Health Care*;**6**:93–103.

Weinstein MC, Manning WG, 1997. Theoretical issues in cost-effectiveness analysis. *J Health Econ*;**16**:121–8.

Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB, for the Panel on Cost-Effectiveness in Health and Medicine, 1996. Recommendations of the panel on cost-effectiveness in health and medicine. *JAMA*;**276**:1253–8.

Weinstein MC, Siegel JE, Garber AM, Lipscomb J, Luce BR, Manning WG, *et al.,* 1997. Productivity costs, time costs and health-related quality of life: a response to the Erasmus group. *Health Econ*;**6**:505–10.

Weissman JS, Levin K, Chasan-Taber S, Massagli MP, Seage GR, Scampini L, 1996. The validity of self-reported health-care utilization by AIDS patients. *AIDS*;**10**:775–83.

West of Scotland Coronary Prevention Study Group, 1995. Computerised record linkage: compared with traditional patient follow-up methods in clinical trials and illustrated in a prospective epidemiological study. *J Clin Epidemiol*;**48**:1441–52.

Whynes D, Walker AR, 1995. On approximations in treatment costing. *Health Econ*;**4**:31–9.

Willan AR, O'Brien BJ, 1996. Confidence intervals for cost-effectiveness ratios: an application of Fieller's theorem. *Health Econ*;**5**:297–305.

Williams A, 1992. Cost-effectiveness analysis: is it ethical? *J Med Ethics*;**18**:7–11.

Williams C, Coyle D, Gray A, Hutton J, Jefferson T, Karlsson G, *et al.,* 1995. European School of Oncology advisory report to the Commission of the European Communities for the 'Europe Against Cancer Programme': cost-effectiveness in cancer care. *Eur J Cancer*;**31A**:1410–24.

Wittes J, Brittain E, 1990. The role of internal pilot studies in increasing the efficiency of clinical trials. *Stat Med*;**9**:65–72.

Wolff N, Helminiak TW, 1996. Nonsampling measurement error in administrative data: implications for economic evaluations. *Health Econ*;**5**:501–12.

Wolff N, Helminiak TW, Kraemer JK, 1997. Getting the cost right in cost-effectiveness analysis. *Am J Psychiatry*;**154**:736–43.

Wonderling D, Langham S, Buxton M, Normand C, McDermott C, 1996a. What can be concluded from the Oxcheck and British family heart studies: commentary on cost effectiveness analyses. *BMJ*;**312**:1274–8.

Wonderling D, McDermott C, Buxton M, Kinmouth AL, Pyke S, Thompson S, *et al.,* 1996b. Costs and cost effectiveness of cardiovascular screening and intervention: the British family heart study. *BMJ*;**312**:1269–73.

Wright P, Lawrence G, 1994. Design of case report forms for clinical trials. Edinburgh: Churchill Livingstone.

Zhou XH, Melfi CA, Hui SL, 1997. Methods for comparison of cost data. *Ann Intern Med*;**127**:752–6.

# Appendix 1

# Methodology of the systematic review

## Overview

The aim of the review was to identify articles in which the methodological issues surrounding costing alongside clinical trials were discussed. The articles were reviewed with the aim of identifying methodological principles and issues relating to the collection and analysis of cost data within the context of clinical trials. The aim was not to record the frequencies with which issues were raised but to be comprehensive in identifying them. In developing the review, structured comment from relevant experts was sought with the aim of detecting points where agreement was lacking and further research required.

## Identifying relevant articles

The process by which relevant articles were identified consisted of ten key stages:

1. definition of inclusion criteria and design of initial search strategy
2. searches of an in-house bibliographic database
3. manual searches of key journals
4. refinement of search strategy
5. electronic searches for key articles
6. review of key papers and identification of key methodological issues
7. electronic searches for articles on specific methodological issues
8. citation searches using articles on specific methodological issues
9. reference lists of identified articles
10. articles identified by experts.

The criteria for the inclusion of articles were devised after consideration of the purpose and scope of the review. As part of the first stage, an initial set of search terms was drawn up and subsequently used to search the in-house database (stage 2). Manual searching of key journals was the third stage. The articles retrieved from the in-house and manual searches were reviewed with the aim of refining the search strategy and defining search terms (stage 4). Electronic searches were conducted using the refined search strategy. The first set of searches was targeted at identifying key review and empirical articles (stage 5). The key articles were then reviewed and a set of key methodological issues identified. These key issues formed the basis of the structured review of the methodological issues (stage 6). A subset of searches was then conducted to identify articles containing discussion of the individual methodological issues (stage 7). A final set of electronic searches was conducted in the form of citation searches to identify key articles (stage 8). The reference lists of articles were also examined to identify further articles (stage 9). The final stage was the identification of articles by experts (stage 10). The ascertainment of methodological issues was an ongoing process and electronic searches for specific issues were revised accordingly.

### Inclusion criteria

The criteria for inclusion of an article in the review were if it:

- reviews methods of costing alongside clinical trials or reviews a single methodological issue
- conducts empirical analysis of economic evaluation alongside a trial that raises new methodological issues
- presents guidelines on how to perform economic evaluation
- conducts empirical analysis of specific methodological issues
- presents guidelines for authors who are publishing economic evaluations in journals.

Although the focus of the review was specific to costing alongside clinical trials, the relevant articles that raised methodological issues were derived from several strands of the literature. First, reviews of the specific topic (costing alongside clinical trials) were included as well as review articles addressing a single methodological issue. Secondly, empirical articles reporting the conducting of economic evaluations alongside or within clinical trials were also included because they might raise additional methodological issues. The relatively recent increase in economic evaluations conducted alongside or within clinical trials meant that many of the methodological issues could be raised in recent empirical

studies. Empirical articles that explored new methodological issues were also included. Thirdly, the literature on costing methods in economic evaluation in general was included because this forms the basis for the design of many economic evaluations. Fourthly, guidelines have recently been produced for various countries, which attempt to standardise methodologies employed for economic evaluation and, because they make recommendations about methods, their inclusion in the review was also appropriate. Finally, guidelines for authors intending to publish economic evaluations in peer reviewed journals were also included in the review, because, by implication, they contain comment on methodological issues.

Abstracts were read and included where inclusion criteria were met. Articles were then skim read and included in the review where inclusion criteria were met. The aim was not to record the frequencies with which issues were raised but to be comprehensive in identifying issues, so the final list of references represents all issues rather than all relevant articles discussing the issues.

## In-house search

The second stage of the review was an in-house search of the relevant literature. The in-house database was organised by researchers in the field of economic evaluation in health care, so it provided a useful starting point for the identification of articles for review, search terms and methodological issues. The in-house database at the Health Economics Research Group, Brunel University, holds over 5000 references organised by using bibliographic software (Pro-cite version 2.21). The basic search terms used for the in-house searches were economic*, cost* or methodology* or trial* (where * indicates a truncated search term). Articles were retrieved when the inclusion criteria were met.

## Manual searches

Manual searching of journals was also conducted. All journals were searched manually from 1990 until October 1997 unless they were first established after 1990. Where this was the case, the first year is presented in parentheses. The following journals were included:

- *British Journal of Medical Economics* (to vol. 11 (1))
- *Controlled Clinical Trials* (to vol. 18 (5))
- *Drug Information Journal* (to vol. 31 (3))
- *Health Economics* (from 1992 to vol. 6 (5))
- *Health Policy* (to vol. 41 (2))
- *International Journal of Technology Assessment in Health Care* (to vol. 13 (3))
- *Journal of Clinical Epidemiology* (to vol. 50 (10))
- *Journal of Epidemiology and Community Health* (to vol. 51 (5))
- *Journal of Health Economics* (to vol. 16 (2))
- *Medical Care* (to vol. 35 (10))
- *Medical Decision Making* (to vol. 17 (3))
- *PharmacoEconomics* (from 1992 to vol. 12 (4)).

## Refinement of search strategy

Following the in-house and manual searches, the articles retrieved were then skim read with a view to refining the search strategy and devising search terms.

In developing the search strategy, use was made of truncated terms (*). Operators such as 'and', 'or' were also used and the operator 'next to' was used for expressions, such as economic evaluation, which became: economic 'next to' evaluation. The use of medical subject headings such as 'costs and costs analysis' and 'economic' were found to be too broad. The use of cost* in abstracts picked up too many articles concerned only with making minor comments about the cost implications of studies. The search terms were combined with their respective operators and further combined into sets:

001 cost* *in title*
002 cost-effective* *in title or abstract*
003 economic *next to* evaluation *in title or abstract*
004 methodology* *in title or abstract*
005 randomised controlled trial* *in title or abstract*
006 1 or 2 or 3
007 5 and 6
008 4 and 7

## Electronic searches

The search strategy detailed above was used for electronic searches, which were conducted simultaneously on MEDLINE, EMBASE and Healthstar (health administration database) via Dialog software. The simultaneous searching of the three databases permitted the easy identification of duplicate articles. Further searches were conducted on the Health Economic Evaluations Database (Office of Health Economics). The literature searches were limited to English language articles for the period 1986–1996 and excluded animal studies.

## Review of key articles and identification of key methodological issues

The articles retrieved from the search were then reviewed in order to identify the methodological issues. The issues were structured into four categories: study design, data collection,

data analysis and presentation of results. These categories also formed the outline of the report.

### Issue-specific searches

The aim was to identify papers discussing a specific methodological issue or issues that are relevant to data collection and analysis for costing alongside clinical trials. This required a targeted search strategy; so, for each key issue, search terms were identified. Electronic searches were conducted for each issue on EMBASE, MEDLINE and Healthstar databases via Dialog.

The issues identified were resource-use data collection and analysis issues (sample size, cost-effectiveness ratios, confidence intervals, sensitivity analysis, extrapolation and generalisation). The search strategy was as follows:

001 statistic* *in title or abstract*
002 stochastic *in title or abstract*
003 cost-effective* *next to* ratio* *in title or abstract*
004 extrapolat* *in title or abstract*
005 sensitivity *next to* analys* *in title or abstract*
006 generali* *next to* cost* *in title or abstract*
007 missing *next to* data *in title or abstract*
008 pool* *next to* data *in title or abstract*
009 cost-effective* *in title or abstract*
010 1 or 2 or 3
011 4 or 5 or 6 or 7 or 8
012 9 and 10
014 9 and 11

These search terms were then combined with those from the general search.

### Citation searches

For each key issue, key articles were selected and used in the citation searches, which were conducted on EMBASE. The articles selected were those thought to be the earliest ones discussing the methodological issues. Identified articles were retrieved if they met the inclusion criteria and were not duplicates.

### Reference lists

All reference lists from previously identified articles were reviewed in order to identify other relevant articles, subject to the inclusion criteria.

### Articles from experts

In the consultation phase of the report, experts were asked whether any key articles were missing from those identified by the searches and, if so, to identify them.

## Review of articles

The articles were photocopied and details entered on Pro-cite (a bibliographic software package) with keywords assigned for the source (in-house, handsearch, electronic search with database) and for the type of article (methodological, empirical, statistical, other review, data collection). The allocation of source keywords enabled easy cross-checking of in-house articles with those retrieved from the electronic searches. For each stage of the identification process, the number of articles identified and the number included were recorded. The results are shown in *Table 2*.

***TABLE 2*** *Number of articles retrieved at each stage*

| Stage | No. articles identified | No. articles included |
|---|---|---|
| In-house search | 174 | 24 |
| Manual search | n/a | 81 |
| Main electronic search | 397 | 76 |
| Issue searches (all) | 143 | 18 |
| Citation searches (all) | 161 | 16 |
| Reference lists | n/a | 24 |
| Experts | n/a | 14 |
| Total | | 253[a] |

[a] *As explained above, because the aim was not to record the frequencies with which issues were raised but to be comprehensive in identifying issues, the final list of references represents all issues rather than all relevant articles discussing the issues*

*n/a, not applicable*

## Experts consulted

Experts were asked to comment on drafts of the report and to identify any gaps in methodological issues and articles in the review. The following experts were consulted:

Martin Backhouse, Sandoz Pharma, Basel, Switzerland
Ray Churnside, Glaxo Wellcome, London, UK
Doug Coyle, Clinical Epidemiology Unit, Ottawa, Canada
Linda Davies, Centre for Health Economics, University of York, UK
Professor Cam Donaldson, Health Economics Research Unit, University of Aberdeen, UK

Dr Miranda Mugford, School of Health Policy and Practice, University of East Anglia, UK
Dr Max Parmar, MRC Cancer Trials Office, Cambridge, UK
Dr Kevin Schulman, Clinical Economics Unit, Georgetown University, USA
Ken Wright, Centre for Health Economics, University of York, UK.

# Appendix 2

# An empirical example of the usefulness of existing data to inform design

## Introduction

The design of many economic evaluations has been based on vague judgements about which costs might be important, and even vaguer ideas of the variability and distribution of such costs (Drummond and O'Brien, 1993; O'Brien *et al.*, 1994; Spiegelhalter *et al.*, 1996). Although this may have been inevitable in the past, there is now some evidence that might usefully inform these decisions. A small number of published articles have begun to show how existing empirical evidence can inform methodological decisions (Knapp and Beecham, 1993; Whynes and Walker, 1995). Given that it is often of further concern that some economic evaluations may overburden a study with detailed data collection when a more restricted set of data could have provided the necessary information, it is important to ensure that cost data collection is designed in a precise and cost-effective way.

This part of the review provides a demonstration of how existing evidence relating to costs from a completed study can be used to inform subsequent design decisions. Analysis of the results of a detailed data collection may show whether (at least in those particular circumstances) simpler methods would have produced comparable results.

## Summary of trial and cost data

This empirical example is based on data from the CHART study. The CHART data were used because of the depth and range of the cost collection. The study is a randomised clinical trial comparing CHART treatment with conventional treatment for patients with head and neck cancer and cancer of the bronchus. The advantages of CHART, compared with conventional treatment, are its potential both to overcome radiation resistance and to increase survival (Saunders *et al.*, 1996[1]).

### Comparators

Patients receiving CHART were treated three times on each of 12 consecutive days, including weekends. The frequency and intensity of treatment meant that they usually stayed in hospital during therapy, either in a hospital ward or in a hostel owned by the hospital. Patients receiving conventional therapy were treated once a day, 5 days per week, for a total of 6 weeks (bronchus cancer) or 6.5 weeks (head and neck cancer). They usually travelled daily to the hospital to receive treatment.

### Design of the trial

The trial was multicentred, with patients recruited from ten UK centres and three centres from other parts of Europe. Sixty per cent of the patients were randomised to receive CHART and 40% conventional treatment. For the purposes of the re-analysis discussed here, only patients with head and neck cancer were included. The total number of patients in the trial of treatment for head and neck cancer was 526, with 314 randomised to receive CHART and 212 randomised to receive conventional treatment.

### Types of cost

These two treatments are likely to have different radiotherapy, hospital and travel costs. One would expect the radiotherapy costs and hospital costs for patients treated by CHART to be higher than those receiving conventional treatment, but the travel costs incurred by patients receiving conventional treatment to be higher than those for patients receiving CHART. Furthermore, if, as a result of more intensive treatment, CHART leads to more side-effects, higher community costs may be incurred by these patients than by those receiving conventional treatment. This re-analysis discusses these cost differences and analyses total costs. There is no discussion of cost-effectiveness because the effects data are not yet available.

[1] Saunders MI, Dische S, Barrett A, Parmar MKB, Harvey A, Gibson D, on behalf of the CHART Steering Committee. Randomised multi centre trials of CHART vs conventional radiotherapy in head and neck and non-small cell lung cancer: an interim report. *Br J Cancer* 1996;**73**:1455–62.

The costs measured in the study followed the usual practice of identifying and measuring resource-use and unit costs separately. Costs were then the product of the resource-use and the unit costs. Five main elements of resource use were measured:

**Radiotherapy resources**
- the number of treatments given
- the timing of the treatments (whether during normal working hours (Monday–Friday), before or after normal hours or at weekends).

**Hospital resource use up to 8 weeks**
- the number of inpatient bed days during and after treatment
- hospital travel time (ambulance).

**Hospital resource use from 8 weeks and to 3 months**
- the number of inpatient bed days by hostel/ward
- the number of outpatient visits
- the number of hostel bed days.

**Community resources**
- the number of surgery visits to the GP for this illness
- the number of surgery visits to the GP for other illnesses
- the number of home visits by the GP for this illness
- the number of home visits by the GP for other illnesses
- the number of visits by district nurses, health visitors, social workers, Macmillan nurses, home helps and other community service workers.

**Patient and hospital travel time**
- the number of visits to the centre per week for treatment by mode of transport
- the distance from the patient's home to the centre.

## Resource-use and unit cost data

The collection of resource-use data was planned to coincide with clinical assessments, the timing of which had already been decided at the start of the trial. Details of radiotherapy resource use were collected at the end of treatment. The use of hospital inpatient and outpatient resources was collected weekly up to 8 weeks after the commencement of treatment. After 8 weeks, this information was collected at 3, 6, 9, 12, 18 and 24 months after commencement of treatment. Information on the use of community resources was collected by asking both the patient and the GP. Details of patients' travel during treatment were collected weekly until treatment was completed. The resource-use questionnaires were piloted at two centres. The resource-use and unit cost measures are summarised in *Table 3*.

***TABLE 3*** *Summary of resource-use and unit cost measurement*

| Resource use | Unit cost |
|---|---|
| ***Radiotherapy*** | Cost per treatment per centre, by length of session |
| ***Hospital*** | |
| Inpatient bed days | Cost per inpatient bed day per centre |
| Hostel bed days | Cost per hostel bed day per centre with hostel |
| Hospital outpatient visits | Cost per outpatient visit per centre |
| Day patient days | Cost per day per centre |
| ***Community*** | |
| GP | Cost per visit (single estimate) |
| District nurse | Cost per visit (single estimate) |
| Health visitor | Cost per visit (single estimate) |
| Social worker | Cost per visit (single estimate) |
| Macmillan nurse | Cost per visit (single estimate) |
| Home help | Cost per visit (single estimate) |
| ***Travel*** | |
| Miles travelled | Cost per mile per mode of travel (single estimate) |

The unit cost estimates (averaged across the centres) are summarised in *Table 4*. The radiotherapy unit costs are costs per patient treatment; the bed day unit costs are cost per bed day; all community unit costs are costs per visit. In calculating total costs, centre-specific resource use was multiplied by centre-specific unit costs where appropriate. The unit costs of radiotherapy and hospital resource use were collected from the individual centres. The unit costs of community resource use were based on previous estimates (Netten and Dennett, 1993[2]). The cost of a GP visit was based on Treasury estimates, modified to allow for differences in the time for surgery and home visits. Total cost up to 3 months after treatment was then the sum of the radiotherapy costs, hospital costs up to 8 weeks, hospital costs from 8 weeks to 3 months, community costs and travel costs.

[2] Netten A, Dennett J. Unit costs of community care. Canterbury: Personal Social Services Research Unit, University of Kent, 1993.

***TABLE 4*** *Summary of unit cost estimates*

| Resource use | Unit cost (UK £ 1993–1994) |
|---|---|
| Radiotherapy (normal hours) | 18.86 |
| Radiotherapy (after normal hours) | 45.52[a] |
| Radiotherapy (weekends) | 53.26[a] |
| Radiotherapy (before normal hours) | 32.37[b] |
| Inpatient bed days | 99.34[a] |
| Hostel bed days | 41.53[b] |
| Hospital outpatient visit | 24.39[a] |
| GP visit | 10.87 |
| District nurse | 9.32 |
| Health visitor | 15.52 |
| Social worker | 18.66 |
| Macmillan nurse | 39.16 |
| Home help | 6.36 |
| Miles travelled (patient travel) | 0.14 to 0.28[c] |

[a] *Mean cost for the ten centres*
[b] *Mean cost for four centres*
[c] *Depends on area*

# Issues in the design of resource-use and cost data collection

## Issues and stages of analysis

The data set described above can be used to explore several issues in the design of resource-use and cost data collection:

- contribution of types of costs to total cost
- variability in cost-generating events and total cost
- sample size
- centre effects.

Together with the methods used to address them, these issues are described in further detail below.

The data are analysed in three stages, as summarised in *Table 5*. The first two stages are based on limited sets of data and aim to reflect the stages that investigators themselves would go through in designing data collection.

**The first stage** is based on a subset of conventionally treated patients only. This data set reflects the likely situation of an investigator who is designing cost data collection, that is, having some prior knowledge of the resource use associated with the existing conventional treatment but no information on the trial intervention. This prior knowledge need not be derived from trials. For example, investigators may have information available to them from patients' notes, a prior descriptive study, or a prior trial of which only one arm was relevant. Data in this first stage are based on 38 patients who had received conventional treatment from one of the larger trial centres. The assumption is that the conventional treatment is the same in routine practice as in the trial.

**The second stage** is based on a simulated pilot study and aims to reflect the next stage that an investigator might go through, collecting data on a small number of patients in each treatment group. The subset of data used in this analysis is based on the first 50 patients randomised over a 3-month period (selected on the basis of date of randomisation) to either treatment in the trial. The number of patients in the pilot study receiving conventional treatment was 23, and 27 received CHART. The simulated pilot study approach has been used by Drummond and Coyle (1997).

**The third stage** is based on the full trial data set, comparing the results for the two treatment groups. The implications that each stage of the analysis has for design issues are compared.

## Contribution of types of cost to total cost

### *Description of the issue*

An examination of the contribution that each type of cost makes to the total cost is useful because it can determine whether a reduced number of costs

***TABLE 5*** *Summary of stages of analysis*

| Stage of analysis | Treatment | Data | Sample size |
|---|---|---|---|
| Stage 1 | Conventional treatment | Existing data | 38 |
| Stage 2 | Both treatment groups | Pilot study | 50 |
| Stage 3 | Both treatment groups | Full trial data set | 526 |

could safely be collected. The relative contribution of each type of cost is assessed by calculating the proportion of total cost accounted for by each type of cost.

The size of contribution to the total cost may be determined by the frequency of the occurrence of the cost and the size of the unit cost. Resource uses with high unit costs may not contribute a large amount to the total cost if they are of low frequency. The relationship between the frequency and level of unit costs of each of the types of cost is therefore explored. A judgement is required about the threshold below which certain types of cost are considered to be of low quantitative importance and, as such, may not require measurement. A previous study has suggested excluding measurement of costs that contribute between 6% and 9% of the total cost (Knapp and Beecham, 1993).

### *Comparison of results by stage of analysis*

*Table 6* presents the five cost types as percentages of the total cost. Stage 1 analysis identifies hospital costs up to 8 weeks, and radiotherapy and travel costs as the three largest components of the total. Hospital costs from 8 weeks to 3 months and community costs contribute a small amount, accounting for 4.8% and 1% of the total respectively.

The stage 2 analysis confirms hospital costs from 8 weeks to 3 months and community costs to be of low quantitative importance for both treatment groups. It also identifies that travel costs are less important for the CHART group than for the conventional group. This would be expected because CHART, unlike conventional treatment, does not require patients to travel daily to the hospital. Stage 3 analysis confirms hospital costs from 8 weeks to 3 months and community costs to be of less quantitative importance than other types of cost.

The stages of analysis have the following implications for the study design. The low contribution of community costs (representing between 1% and 4% of the total) and the fact that they are of low frequency and have low unit costs suggests that, if detailed data collection on community visits were not collected, little information would be lost. The relatively large contribution of travel costs for conventionally treated patients confirms their importance for this group. The decision about whether travel costs are to be included will, ultimately, depend on the perspective of the study. Hospital costs from 8 weeks to 3 months also represent a small proportion of the total cost for both treatment groups at all stages of the analysis; this suggests that the additional information

***TABLE 6*** *Cost type as percentage of total cost*

| Cost type | Stage of analysis | Conventional | | CHART | |
|---|---|---|---|---|---|
| | | % of total cost | Mean cost (£) | % of total cost | Mean cost (£) |
| Radiotherapy costs | Conventional | 33.2 | 747.38 | n/a | n/a |
| | Pilot | 22.7 | 549.74 | 36.2 | 1246.40 |
| | Actual | 25.3 | 587.01 | 34.3 | 1171.10 |
| Hospital costs to 8 weeks | Conventional | 43.6 | 980.20 | n/a | n/a |
| | Pilot | 42.7 | 1035.53 | 56.2 | 1934.90 |
| | Actual | 45.0 | 1044.74 | 57.8 | 1975.41 |
| Hospital costs 8 weeks to 3 months | Conventional | 4.8 | 108.67 | n/a | n/a |
| | Pilot | 6.0 | 145.74 | 3.4 | 118.26 |
| | Actual | 8.2 | 189.72 | 4.8 | 163.11 |
| Community costs | Conventional | 1.0 | 23.04 | n/a | n/a |
| | Pilot | 2.2 | 52.52 | 3.4 | 118.27 |
| | Actual | 4.0 | 93.13 | 2.5 | 84.16 |
| Travel costs | Conventional | 17.4 | 391.37 | n/a | n/a |
| | Pilot | 26.4 | 642.20 | 0.7 | 23.74 |
| | Actual | 17.5 | 407.96 | 0.6 | 20.89 |
| Total | Conventional | 100 | 2250.66 | n/a | n/a |
| | Pilot | 100 | 2425.73 | 100 | 3441.57 |
| | Actual | 100 | 2322.54 | 100 | 3414.67 |

*n/a, not applicable*

acquired by extending the data collection by 1 month may be low. When the frequency and level of unit costs are considered, however, it may be important to include hospital costs from 8 weeks to 3 months because, although they are of low frequency, they have high unit costs.

Had stage 1 and stage 2 analyses been performed, the design of the main study would have excluded community costs and, possibly, hospital costs from 8 weeks to 3 months. These design decisions are confirmed by the stage 3 analysis of the full data, which, on the basis of their contribution to total cost, collected additional data that would not have been necessary.

The decision on types of cost to be included should also address variability in costs, which will now be discussed.

## Variability in cost-generating events and total cost

### *Description of the issue*

A knowledge of the variability in cost-generating events and cost data is important in informing the types of cost to be included in a study and in sample size calculations. This variability is explored by examining the dispersion of the events. Dispersion refers to the ratio between the variance and the mean of a variable. If the variance of a cost-generating event is greater than its mean, it is said to be overdispersed. If the variance equals the mean, there is no overdispersion; therefore, the higher the ratio, the more overdispersed the data. This measurement has been applied by Spiegelhalter and colleagues (1996), who suggest that a typical range of overdispersion for cost-generating events is between 2 and 5.

The variability in total cost is examined by calculating the coefficient of variation, which expresses the standard deviation of a cost-generating event as a percentage of the mean. This allows the variability between different types of cost and total cost for the treatment groups to be compared.

### *Comparison of results by stage of analysis*

*Table 7* presents the variability in cost-generating events and illustrates that overdispersion varies by type of cost-generating event and between

***TABLE 7*** *Variability in cost-generating events (variance/mean ratio)*

| Cost-generating event | Stage | Conventional | CHART |
|---|---|---|---|
| Number of radiotherapy treatments | Conventional | –[a] | n/a |
| | Pilot | –[a] | –[a] |
| | Actual | 0.002 | 0.15 |
| Days immediately post-treatment | Conventional | 7.5 | n/a |
| | Pilot | 9.4 | 20.0 |
| | Actual | 8.3 | 12.9 |
| Days in hostel during treatment | Conventional | 1.0 | n/a |
| | Pilot | 27.0 | 8.9 |
| | Actual | 30.8 | 11.2 |
| Days in ward during treatment | Conventional | 33.6 | n/a |
| | Pilot | 23.1 | 3.7 |
| | Actual | 26.6 | 4.5 |
| Inpatient days to 3 months | Conventional | 26.2 | n/a |
| | Pilot | 4.6 | 9.4 |
| | Actual | 23.9 | 19.0 |
| Outpatient days to 3 months | Conventional | 0.3 | n/a |
| | Pilot | 1.0 | 1.1 |
| | Actual | 1.1 | 1.2 |
| Community visits (all) | Conventional | 23.9 | n/a |
| | Pilot | 12.1 | 64.4 |
| | Actual | 87.2 | 42.2 |

[a] *Zero variance observed, therefore no ratio calculated*

*n/a, not applicable*

treatment groups for the same event. Compared with their means, for all three stages of the analysis, there is low variability in the number of radiotherapy treatments. This would be expected, given that all patients undergo radiotherapy, the number of treatments being predetermined in the therapy or trial protocol. This is not the case for other cost-generating events, such as the number of days spent on the ward immediately after treatment, which is not set in the trial protocol and hence may vary more. Compared with their means, a lower variation in the number of days spent on the ward immediately after treatment is observed for the conventional treatment group compared with the CHART group. For the stages 2 and 3 analyses, compared with their means, a lower variation in the number of days on the ward during treatment is observed for CHART than for conventional treatment. Again, this is not surprising because days spent on the ward during treatment formed part of the CHART protocol. Comparing the results of the stage 2 analysis with those of stage 3, no pattern emerges of whether, for example, the stage analysis (pilot study) over- or underestimated actual overdispersion. This may be contrary to expectation because one might presume pilot studies to display less overdispersion because they are based on fewer observations.

The implications for the study design are that data should be collected on cost-generating events with high variability and that data need not be collected on those with low variability. Furthermore, if the number of a particular cost-generating event is specified in the protocol, then it can be treated as fixed and detailed data need not be collected on it.

*Table 8* presents the variability of total costs for the two treatment groups. For all stages of analysis, compared with their means, radiotherapy costs are the least variable type of cost. There is a higher variability in radiotherapy costs for CHART than for the conventional treatment group. This may have arisen because of the higher unit costs attached to radiotherapy treatments taking place out of normal working hours, which were more likely to occur in the CHART group. For the stages 2 and 3 analyses, compared with their means, the total cost of conventional treatment varies more than that of CHART. This may suggest that conventional treatment was less specified in the treatment protocol and that more practice variation existed between centres in terms of the conventional treatment provided. When comparing the stages 2 and 3 analyses, there is no consistent pattern in terms of whether the stage 2 analysis (pilot study) over- or underestimates variability in the type of cost and the total cost.

***TABLE 8*** *Variability in total costs*

| Cost type | Stage | Coefficient of variation (%) | |
|---|---|---|---|
| | | Conventional | CHART |
| Radiotherapy costs | Conventional | 0.5 | n/a |
| | Pilot | 20.9 | 42.4 |
| | Actual | 20.4 | 34.7 |
| Hospital costs to 8 weeks | Conventional | 195.9 | n/a |
| | Pilot | 130.7 | 77.8 |
| | Actual | 334.4 | 310.8 |
| Hospital costs 8 weeks to 3 months | Conventional | 463.2 | n/a |
| | Pilot | 130.7 | 253.3 |
| | Actual | 156.0 | 63.2 |
| Community costs | Conventional | 323.8 | n/a |
| | Pilot | 165.6 | 218.6 |
| | Actual | 423.7 | 232.4 |
| Travel costs | Conventional | 158.8 | n/a |
| | Pilot | 227.7 | 167.5 |
| | Actual | 195.7 | 282.6 |
| Total cost | Conventional | 96.3 | n/a |
| | Pilot | 87.0 | 59.8 |
| | Actual | 91.6 | 49.7 |

*n/a, not applicable*

An important issue to address is the relationship between the cost-generating events and their associated unit costs. As all stages of the analysis show, even though there may be low variation in the number of cost-generating events (in this case, the number of radiotherapy treatments), if the unit costs differ according to when the event occurs then this may give rise to additional variation in the particular type of cost. As with the previous analysis on variability in cost-generating events, data on types of cost with low variability need not be collected.

## Sample size

### *Description of the issue*

In calculating sample size, assumptions have to be made about:

- the underlying distribution of the data
- the desired level of significance
- the desired level of power
- the minimum difference that the study will be able to detect.

Information on the variability in cost data is also required. The sample size for the economic evaluation was not formally addressed in the study design for the economic evaluation of CHART. The sample size that would have been required to test for differences in total cost, and the difference in total cost that can be detected from the cost data, was assessed by performing retrospective power calculations. This approach has been used in another study (Gray *et al.*, 1997). Different perspectives may imply different sample sizes (Drummond and O'Brien, 1993), so the sample size requirements to test for difference in total cost and NHS cost are also examined. The NHS cost is the sum of the radiotherapy, hospital and community costs; the total cost is the NHS cost plus travel costs.

The assumptions made in the sample size calculations are as follows:

- a normal distribution
- 5% significant level
- both 80% and 90% power are used
- a 20% difference in total cost is assumed to be worth detecting.

Variability in the data is based on actual standard deviations. Because the trial was designed with a study group 1.5 times the size of the control group, the same design is assumed and thus sample size calculations are based on unequal sample sizes per group. This allows for comparison between the actual study sample size and those estimated by the various stages of analysis. Sample sizes are determined for the stage 2 and stage 3 analyses. The difference in total cost that the study had the power to detect is determined for the stage 3 analysis only.

### *Comparison of results by stage of analysis*

The sample sizes required to detect differences in total cost and NHS cost by treatment are shown in *Table 9*. For the stage 2 analysis, the sample sizes are greater than those indicated for stage 3. Compared

***TABLE 9*** *Sample sizes for total NHS cost and total cost*[a]

| Assumptions | Stage | Sample size required | |
|---|---|---|---|
| | | Conventional | CHART |
| 90% power<br>20% difference in NHS cost | Conventional | n/a | n/a |
| | Pilot | 241 | 362 |
| | Actual | 207 | 311 |
| 80% power<br>20% difference in NHS cost | Conventional | n/a | n/a |
| | Pilot | 180 | 270 |
| | Actual | 155 | 232 |
| 90% power<br>20% difference in total cost | Conventional | n/a | n/a |
| | Pilot | 234 | 335 |
| | Actual | 190 | 285 |
| 80% power<br>20% difference in total cost | Conventional | n/a | n/a |
| | Pilot | 167 | 250 |
| | Actual | 142 | 212 |

[a] *Sample size calculations assume 5% significance level*

*n/a, not applicable*

with the actual sample sizes (314 for CHART and 212 for the conventional treatment group), the stage 2 analysis (pilot study) thus overestimates the sample size required. For the stage 3 analysis, when compared with the actual sample sizes, the required sample sizes are very similar.

The implications for the design of the study are that, if the desired power was reduced to 80%, then the sample size for the CHART group could have been reduced by 82 patients. Thus, the sample size for the economic evaluation could have been smaller than that of the clinical evaluation. This is contrary to the findings of another study, which found that large sample sizes were required to detect cost difference as opposed to differences in clinical outcomes (Gray *et al.*, 1997). A smaller sample size is indicated for detecting differences between total cost and NHS costs.

*Table 10* shows the differences between total cost and NHS cost from the stage 3 analysis (the difference the actual study had the power to detect). It shows that the study had the power to detect differences of approximately 16–20% in cost between treatment groups. The study was therefore highly powered. Whether these cost variations are large enough to be important depends on the policy makers and the difference in benefits derived.

***TABLE 10*** *Detectable differences in cost*

| Assumptions | Difference (£) detectable (%) |
|---|---|
| 90% power (NHS cost) | 572.51 (20.5) |
| 80% power (NHS cost) | 494.45 (17.7) |
| 90% power (total cost) | 562.22 (20.1) |
| 80% power (total cost) | 485.57 (16.3) |

## Unit costs from all centres

In the CHART study, all radiotherapy and hospital resource use was valued using centre-specific unit costs from ten centres. It is useful to explore the extent to which unit cost data collection from all the centres was necessary because the collection of unit cost information is often a time-consuming task involving both the investigators and hospital staff. There is disagreement about whether or not the use of a single set of unit cost estimates conceals resource-use differences. In order to address this issue, a single set of unit costs was used to value resource use and the resulting difference in NHS cost was calculated. The difference in NHS cost, rather than the total cost, was calculated because the unit costs estimates affect the NHS cost only. The single set of unit costs chosen was for one of the larger centres in the trial. This was a somewhat arbitrary selection but, as the purpose was for illustration only, the selection of a centre was not of primary concern. *Table 11* summarises the single set of unit costs used in the analysis. When compared with the average unit costs from all centres presented in *Table 4*, the single set of unit costs is a mixture of higher and lower unit costs.

*Table 12* presents the results from applying a single set of unit costs to centre-specific resource use. The use of a single set of unit costs increases

***TABLE 11*** *Single set of unit costs estimates*

| Resource | Unit cost (£) |
|---|---|
| Radiotherapy in morning session | 22.63 |
| Radiotherapy in daytime session | 22.63 |
| Radiotherapy in weekend session | 58.02 |
| Day on ward | 111.52 |
| Outpatient visit | 17.26 |
| Day in hostel | 41.53 |

***TABLE 12*** *NHS costs arising from use of a single set of unit costs*

| Cost | CHART | | Conventional | |
|---|---|---|---|---|
| | £ Mean (SD) | % change[a] | £ Mean (SD) | % change[a] |
| NHS cost | 3581.30 (1699) | +5.5% | 2302 (2314) | +20.2% |

[a] *% change when compared with NHS using centre-specific resource-use and unit costs*

*SD, standard deviation*

the NHS costs of both treatment groups but increases the costs of the conventionally treated group more. This suggests that, in this case, the use of a single set of unit costs overestimates costs and introduces additional uncertainty. Centre effects are complex but, at the very least, this simple analysis demonstrates that the decision about whether to collect unit costs from all centres can affect the results. The collection of a single set of unit costs estimates may conceal important cost differences and the results may be misleading. Furthermore, if it is the case that unit cost estimates are a function of resource use, then centre-specific unit costs should be collected.

## Implications

This part of the review has demonstrated how existing empirical evidence can inform methodological decisions in the design of a costing study alongside a clinical trial. The three stages of analysis adopted aimed to reflect the stages of knowledge and analysis that investigators would themselves go through.

The stage 1 analysis presented here, based on data likely to be available to investigators, can be a useful first stage in developing a study design and can be performed before any new data collection is carried out. It reflects the situation that an investigator who is designing a cost data collection study is likely to be in, that is, having some prior knowledge on the resource use associated with the existing conventional treatment, from, for example, patient notes or previous studies. If additional data collection is required, a pilot study can be performed. The data from the pilot study could also be used as part of a two-stage design (Wittes and Brittain, 1990), whereby the pilot study data are used in the main analysis.

This re-analysis demonstrates that it is possible to reduce the data collection effort by considering the contribution of types of cost and their variability. In doing so, the importance of thinking through the relationship between the frequency of cost-generating events and their associated unit costs is also identified. It may be possible to omit the collection of data on low-frequency, low-unit cost events. Furthermore, this decision may depend on existing knowledge of the variability in cost-generating events and the extent to which the number of events are specified in the protocol or are fixed for each individual in the trial. The re-analysis has also demonstrated that it is not necessarily the case that pilot studies underestimate variability or that costing studies require larger sample sizes than a clinical study. Finally, it has identified that collecting unit costs from all centres in a multicentre trial is important in order not to conceal cost differences.

Despite the fact that the re-analysis reported here is relatively straightforward, it has identified the usefulness of having access to and examining existing data so that costing studies can be designed in a precise and cost-effective way.

# Appendix 3

## Designing cost data collection alongside clinical trials: a decision aid

The authors of several studies have discussed the methodological issues in conducting economic evaluations alongside clinical trials (Drummond and Stoddart, 1984; Bulpitt *et al.*, 1990a,b; Drummond and Davies, 1991; Adams *et al.*, 1992; Bennett *et al.*, 1994a,b; Drummond, 1994, 1995). Despite the existence of these studies, however, practical advice on how to deal with the methodological issues is not readily available. The main part of this report reviews methodological issues in costing. Many of the decisions about how best to handle these remain unresolved, require further empirical testing, or appear to be based on unwritten rules. In order to bridge the perceived gap between existing methodological statements and practical advice, a logical framework within which decisions can be made about costing alongside clinical trials has been developed and is presented here. It aims to provide those who are designing clinical trials (or reviewing designs) with a structured framework to assist in making appropriate choices relating to the collection of data for costing, given the context of the study, the planned clinical end-points, the particular economic questions posed, etc. Effectively, this involves a set of explicit questions, which address the issues of what costs to measure, how to measure them, from whom to collect data, when to collect the data, and with what instruments, given the particular context of the study and the precise economic question.

The specific resource-use items on which data should be obtained will not prove to be the same for all studies: a 'standardised bolt-on package' will not be applicable to all detailed decisions about costing. It is desirable, however, that consistent logic should be applied, so that arbitrary differences do not arise to compromise the comparability of the results. The decision aid described here aims to provide a common approach in the process of discussing which cost data should, and could, appropriately be included within a clinical trial.

The starting point of the decision aid is that an economic evaluation is required alongside a trial. Important issues have to be addressed in deciding whether a trial is warranted in the first place but, in order to put boundaries round the exercise, this issue is not considered here. The aims of the decision aid are summarised below:

- to address key questions in designing resource-use data collection for costing alongside clinical trials
- to present these questions in a logical sequence
- to indicate the basis upon which these questions can be answered
- to refer the reader to more detailed discussion in the main text.

The processes by which the decision aid was developed and its structure are summarised below:

**Development**
- based on evidence from a systematic review of the literature
- structured at a brainstorming session with invited experts.

**Structure**
- overall structure of the decision aid presented in the form of a flowchart, which emphasises and highlights the interrelationship between the stages of decision making
- flowchart provides a first route through study design decisions that have to be made, but recognises the iterative nature of these decisions and hence allows for interrelationships
- remainder of decision aid presented as summary boxes containing detailed checklists on each key design question in the flowchart, with cross-references to the main text as appropriate.

Ultimately, the order of questions is of less importance than that the checklists are used iteratively so that early decisions are revised as necessary.

The flowchart of key design questions is presented in *Figure 1*. It has ten key design decisions and can be divided into three phases: (1) assessing the task; (2) designing data collection; and (3) testing and iteration. The detailed checklists of questions are presented in summary boxes after the flowchart. The design questions are cross-referenced to the relevant sections in the main text of the review.

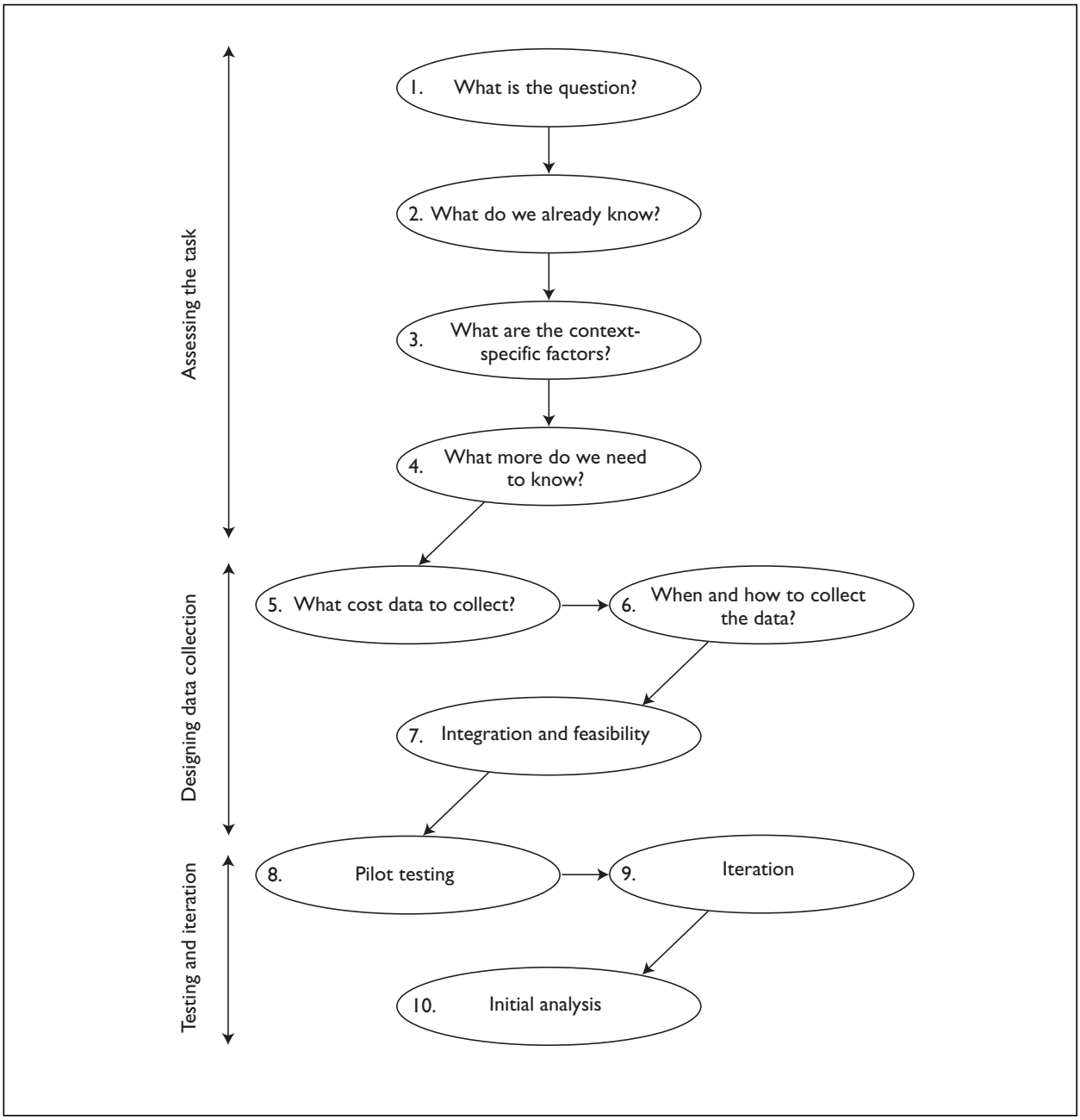



***FIGURE 1*** *Flowchart of key design questions*

## Assessing the task

### 1. What is the question?

*Box 23* presents the checklist for assessing the question. First, the economic question being addressed and its relative importance in economic terms can be defined in terms of the absolute and relative sizes of benefit gained for a given change in cost. This will depend on both the potential numbers involved in the trial and the potential cost and effect differences between the alternatives. Second is a reminder of why the economic evaluation alongside a trial is being performed in the first place, whether it is for policy purposes, who requested the study and so on.

**BOX 23 Checklist for assessing the question**

- What is the economic question?
- Why is the economic evaluation being performed alongside a trial?
- Is the approach hypothesis testing/generating or estimation?
- Is the approach based on welfare economic theory?

*See chapter 2:*
*Economic welfare theory (p. 10)*
*Perspective (p. 10)*
*Hypothesis testing or estimation (p. 15)*

Thirdly, whether an hypothesis-testing or estimation approach is being adopted must be considered. This will also include the possibility that the trial, and the associated economic question, may be hypothesis generating or hypothesis testing. Fourthly, it must be determined whether the preferred approach is based on welfare economic theory or is from a decision-making perspective, that is, adopting a more pragmatic approach to study design. An approach based on welfare economic theory would collect data on all types of cost that the theory deemed relevant. A decision-making approach would determine a set of predefined costs or cost-generating events of interest to the decision maker and limit data collection to these.

## 2. What do we already know?

*Box 24* presents the checklist of questions for assessing existing knowledge. The emphasis is on utilising previous studies and available data sets in order to establish baseline information about the interventions being compared and their costs. The first point in the checklist is concerned with an adequate description of the interventions being compared in the trial. If the interventions and their potential consequences can be described, then identifying resource-use patterns will be made easier. This will allow identification of the relevant economic studies in the clinical area or economic studies addressing similar economic questions. Reviewing previous studies may reveal the key cost-generating events or identify existing data sets that may be useful to access.

**BOX 24 Checklist for assessing existing knowledge**

- Are the trial interventions adequately understood?
- What previous economic studies have been performed?
- What are the main cost-generating events?
- What data are available on the main cost-generating events?
- What data are available to assess variability of cost data between patients?

*See chapter 2:*
*Identifying key cost-generating events (pp. 12–13)*

## 3. What are the context-specific factors?

The trial design itself and other context-specific factors affect the design of the costing study. It is therefore useful to identify the context-specific factors that are likely to influence study design. These constraints are summarised in *Box 25*. Research funds are limited and this may affect both the level of detail of resource-use data collected and the data collection methods. The length of follow-up period in the trial may also affect the methods adopted. A limited time horizon may suggest that a modelling approach is required to supplement data. The primary end-point(s) of the trial may not be 'final' and may be surrogate or composite measures. If the end-point is not a final outcome measure then, again, some form of supplementary modelling may be required. A further constraint may arise if the design of the trial has already been finalised before the economic evaluation has been considered. Finally, statistical considerations, particularly whether the trial is powered to detect cost differences and whether it is multicentred, may be additional context-specific factors.

**BOX 25 Checklist for assessing context-specific factors**

- What are the budget constraints and enrolment costs?
- What is the time horizon of the trial?
- What is the trial end-point?
- Was the trial designed before the design of the costing study?
- What are the statistical considerations?
- Is it a multicentre trial?

*See chapters 2 and 4:*
*Data required from outside the trial (pp. 18–19)*
*Sample sizes (pp. 15–16)*
*Centre selection (pp. 17–18)*
*Modelling long-term outcomes (pp. 19, 32–33)*

## 4. What more do we need to know?

Before deciding on the precise resource-use data to be collected, additional information may be required to enable final design decisions to be made *(Box 26)*. Prestudy data collection methods could be used. For example, a sample of patients external to the trial could be surveyed to gain additional information. This is distinct from a pilot study, which would essentially test the methods. Interviews with key staff involved, such as health-

**BOX 26 Checklist of assessing additional information required to inform design**

- Would a period of prestudy data collection be useful?
- Would interviews with the key staff involved be useful?

*See chapter 2:*
*Identifying key cost-generating events (pp. 12–13)*
*Sampling (p. 17)*

care managers, could be performed to identify sources of data and service organisation.

## Designing data collection

### 5. What cost data to collect?

The decision on the resource-use items on which to collect data depends on several factors. The checklist of questions for this design question are presented in *Boxes 27–29*. *Box 27* concerns the decisions on which costs are to be included; *Box 28* lists the methods for determining the key cost-generating events; and *Box 29* covers decisions about other data requirements.

**BOX 27 Checklist for deciding which types of cost to include**

**Collect resource use for which health care and non-health service resources?**

*Consider:*
- economic welfare theory
- perspective
- form of economic evaluation
- double counting
- quantitative importance
- attribution
- time horizon.

*See chapter 2:*
*Potential factors influencing the types of cost included (pp. 9–12)*

**BOX 28 Checklist for deciding what the key cost-generating events are**

**How to define the key cost-generating events?**

*Consider:*
- variation between arms
- variation between patients within arms
- impact on cost/cost-effectiveness
- consequences if not collected
- hypotheses about events.

**How to identify the key cost-generating events?**

*Consider:*
- reviewing previous studies
- pretrial data collection
- pilot testing
- modelling
- expert opinion
- marginal value of information.

**Level of measurement of resource quantities?**

*Consider:*
- availability of unit cost data.

*See chapter 2:*
*Identifying key cost-generating events (pp. 12–13)*
*Relating resource-use measurement to valuation (p. 14)*

**BOX 29 Checklist for deciding other data requirements**

**Collect data on routine practice?**

*Consider:*
- generalisability of results.

**Collect data on patients' demographic characteristics?**

*Consider:*
- relationship of demographic characteristics to clinical outcome.

**Collect information on patients prior to randomisation?**

*Consider:*
- prerandomisation outcomes as postrandomisation outcomes.

*See chapter 2:*
*Data required from outside the trial (pp. 18–19)*

### 6. When and how to collect data?

How and when the data should be collected is the sixth design question, the checklists for which are presented in *Boxes 30* and *31*. *Box 30* concerns sampling strategy, and *Box 31* data collection methods.

### 7. Integration and feasibility

The seventh design question concerns reviewing the design decision to see if it can be integrated as a coherent whole and ensuring that all data collected relate to the study objective and economic question. The checklist for integrating and assessing feasibility is presented in *Box 32*.

**BOX 30 Checklist for determining sampling strategy**

**Collect resource use for all patients?**

*Consider:*
- variability between patients
- patient burden.

**Collect resource use from patients outside the trial?**

*Consider:*
- comparability of patients.

**Collect resource use from all centres?**

*Consider:*
- centre characteristics.

**Level of power and precision?**

*Consider:*
- type I and II errors.

*See chapter 2:*
*Sampling strategy (pp. 15–17)*

**BOX 31 Checklist for assessing how and when to collect resource-use data**

**Collect patient-specific data directly from patients by questionnaire, interview, diary cards?**

*Consider:*

- forms of bias
- response rates
- patient burden.

**Collect patient-specific data from existing records?**

*Consider:*

- access to records
- retrievability of records
- measurement error.

**When to collect resource-use data?**

*Consider:*

- time cycle of disease
- trial time horizon
- protocol time points.

**How to organise data collection?**

*Consider:*

- data monitoring
- quality assurance.

*See chapters 2 and 3:*
*Timing and frequency of resource-use measurement (pp. 14–15*
*Patient-specific resource use from patients (pp. 21–23)*
*Patient-specific resource use from questionnaires and case record forms completed by others (pp. 23–24)*
*Patient-specific resource use from existing records and databases (pp. 24–25)*
*Mixed methods of data collection (p. 26)*
*Organisation of data collection (p. 26)*

**BOX 32 Checklist for assessing integration and feasibility**

- Is the plan reasonable?
- Is the plan feasible with respect to the constraints?
- Is the burden on participants, clinicians and staff acceptable?
- Do all data collected relate to the study objective and economic question?

## Testing and iteration

### 8. Pilot testing

The eighth key design stage involves the pilot testing of methods, for example, the piloting of questionnaires to be used to measure patients' resource use. The aim of this is not to produce data but to test the methods adopted. The data collection methods may then need to be revised in the light of the pilot testing. A two-stage design could be adopted; for example, internal pilot studies could be performed where the data collected are used in later analyses (Wittes and Brittain, 1990).

### 9. Iteration

The ninth stage is to review all the decisions taken in the light of previous decisions.

### 10. Initial analysis to inform the rest of the study

The final key design stage is an initial analysis of results. Part of this process will involve data monitoring. If a two-stage design has been adopted, then data from the internal pilot study can be used. Changes made at this stage would be simplifications of rather than additions to data collection. For example, data collection on a particular cost-generating event could be stopped if there was little variation or a low level of resource use.

# Health Technology Assessment panel membership

This report was identified as a priority by the Methodology Panel.

## Acute Sector Panel

**Current members**

**Chair:**
**Professor Francis H Creed**, University of Manchester

Professor Clifford Bailey, University of Leeds

Ms Tracy Bury, Chartered Society of Physiotherapy

Professor Collette Clifford, University of Birmingham

Dr Katherine Darton, M.I.N.D.

Mr John Dunning, Papworth Hospital, Cambridge

Mr Jonathan Earnshaw, Gloucester Royal Hospital

Mr Leonard Fenwick, Freeman Group of Hospitals, Newcastle-upon-Tyne

Professor David Field, Leicester Royal Infirmary

Ms Grace Gibbs, West Middlesex University Hospital NHS Trust

Dr Neville Goodman, Southmead Hospital Services Trust, Bristol

Professor Mark P Haggard, MRC

Professor Robert Hawkins, University of Manchester

Dr Duncan Keeley, General Practitioner, Thame

Dr Rajan Madhok, East Riding Health Authority

Dr John Pounsford, Frenchay Hospital, Bristol

Dr Mark Sculpher, University of York

Dr Iqbal Sram, NHS Executive, North West Region

**Past members**

Professor John Farndon, University of Bristol*

Professor Senga Bond, University of Newcastle-upon-Tyne

Professor Ian Cameron, Southeast Thames Regional Health Authority

Ms Lynne Clemence, Mid-Kent Health Care Trust

Professor Cam Donaldson, University of Aberdeen

Professor Richard Ellis, St James's University Hospital, Leeds

Mr Ian Hammond, Bedford & Shires Health & Care NHS Trust

Professor Adrian Harris, Churchill Hospital, Oxford

Dr Gwyneth Lewis, Department of Health

Mrs Wilma MacPherson, St Thomas's & Guy's Hospitals, London

Dr Chris McCall, General Practitioner, Dorset

Professor Alan McGregor, St Thomas's Hospital, London

Professor Jon Nicholl, University of Sheffield

Professor John Norman, University of Southampton

Professor Michael Sheppard, Queen Elizabeth Hospital, Birmingham

Professor Gordon Stirrat, St Michael's Hospital, Bristol

Dr William Tarnow-Mordi, University of Dundee

Professor Kenneth Taylor, Hammersmith Hospital, London

## Diagnostics and Imaging Panel

**Current members**

**Chair:**
**Professor Mike Smith**, University of Leeds

Dr Philip J Ayres, Leeds Teaching Hospitals NHS Trust

Dr Paul Collinson, Mayday University Hospital, Thornton Heath

Dr Barry Cookson, Public Health Laboratory Service, Colindale

Professor David C Cumberland, University of Sheffield

Professor Adrian Dixon, University of Cambridge

Mr Steve Ebdon-Jackson, Department of Health

Mrs Maggie Fitchett, Association of Cytogeneticists, Oxford

Dr Peter Howlett, Portsmouth Hospitals NHS Trust

Professor Alistair McGuire, City University, London

Dr Andrew Moore, Editor, *Bandolier*

Dr Peter Moore, Science Writer, Ashtead

Professor Chris Price, London Hospital Medical School

Dr William Rosenberg, University of Southampton

Dr Gillian Vivian, Royal Cornwall Hospitals Trust

Dr Greg Warner, General Practitioner, Hampshire

**Past members**

Professor Michael Maisey, Guy's & St Thomas's Hospitals, London*

Professor Andrew Adam, Guy's, King's & St Thomas's School of Medicine & Dentistry, London

Dr Pat Cooke, RDRD, Trent Regional Health Authority

Ms Julia Davison, St Bartholomew's Hospital, London

Professor MA Ferguson-Smith, University of Cambridge

Dr Mansel Hacney, University of Manchester

Professor Sean Hilton, St George's Hospital Medical School, London

Mr John Hutton, MEDTAP International Inc., London

Professor Donald Jeffries, St Bartholomew's Hospital, London

Dr Ian Reynolds, Nottingham Health Authority

Professor Colin Roberts, University of Wales College of Medicine

Miss Annette Sergeant, Chase Farm Hospital, Enfield

Professor John Stuart, University of Birmingham

Dr Ala Szczepura, University of Warwick

Mr Stephen Thornton, Cambridge & Huntingdon Health Commission

Dr Jo Walsworth-Bell, South Staffordshire Health Authority

* Previous Chair

*continued*

*continued*

## Methodology Panel

**Current members**

**Chair:**
**Professor Martin Buxton**, Brunel University

Professor Doug Altman, Institute of Health Sciences, Oxford

Dr David Armstrong, Guy's, King's & St Thomas's School of Medicine & Dentistry, London

Professor Nick Black, London School of Hygiene & Tropical Medicine

Professor Ann Bowling, University College London Medical School

Dr Mike Clarke, University of Oxford

Professor Michael Drummond, University of York

Dr Vikki Entwistle, University of Aberdeen

Professor Ewan Ferlie, Imperial College, London

Professor Ray Fitzpatrick, University of Oxford

Professor Jeremy Grimshaw, University of Aberdeen

Dr Stephen Harrison, University of Leeds

Mr John Henderson, Department of Health

Professor Richard Lilford, Regional Director, R&D, West Midlands

Professor Theresa Marteau, Guy's, King's & St Thomas's School of Medicine & Dentistry, London

Dr Henry McQuay, University of Oxford

Dr Nick Payne, University of Sheffield

Professor Margaret Pearson, NHS Executive North West

Professor David Sackett, Centre for Evidence Based Medicine, Oxford

Dr PAG Sandercock, University of Edinburgh

Dr David Spiegelhalter, Institute of Public Health, Cambridge

Professor Joy Townsend, University of Hertfordshire

**Past members**

Professor Anthony Culyer, University of York *

Professor Michael Baum, Royal Marsden Hospital

Dr Rory Collins, University of Oxford

Professor George Davey-Smith, University of Bristol

Professor Stephen Frankel, University of Bristol

Mr Philip Hewitson, Leeds FHSA

Mr Nick Mays, King's Fund, London

Professor Ian Russell, University of York

Dr Maurice Slevin, St Bartholomew's Hospital, London

Professor Charles Warlow, Western General Hospital, Edinburgh

## Pharmaceutical Panel

**Current members**

**Chair:**
**Professor Tom Walley**, University of Liverpool

Dr Felicity Gabbay, Transcrip Ltd

Mr Peter Golightly, Leicester Royal Infirmary

Dr Alastair Gray, Health Economics Research Unit, University of Oxford

Professor Rod Griffiths, NHS Executive West Midlands

Mrs Jeanette Howe, Department of Health

Professor Trevor Jones, ABPI, London

Ms Sally Knight, Lister Hospital, Stevenage

Dr Andrew Mortimore, Southampton & SW Hants Health Authority

Mr Nigel Offen, Essex Rivers Healthcare, Colchester

Mrs Marianne Rigge, The College of Health, London

Mr Simon Robbins, Camden & Islington Health Authority, London

Dr Frances Rotblat, Medicines Control Agency

Dr Eamonn Sheridan, St James's University Hospital, Leeds

Mrs Katrina Simister, Liverpool Health Authority

Dr Ross Taylor, University of Aberdeen

**Past members**

Professor Michael Rawlins, University of Newcastle-upon-Tyne*

Dr Colin Bradley, University of Birmingham

Professor Alasdair Breckenridge, RDRD, Northwest Regional Health Authority

Ms Christine Clark, Hope Hospital, Salford

Mrs Julie Dent, Ealing, Hammersmith & Hounslow Health Authority, London

Mr Barrie Dowdeswell, Royal Victoria Infirmary, Newcastle-upon-Tyne

Dr Tim Elliott, Department of Health

Dr Desmond Fitzgerald, Mere, Bucklow Hill, Cheshire

Professor Keith Gull, University of Manchester

Dr Keith Jones, Medicines Control Agency

Dr John Posnett, University of York

Dr Tim van Zwanenberg, Northern Regional Health Authority

Dr Kent Woods, RDRD, Trent RO, Sheffield



* Previous Chair

# Population Screening Panel

**Current members**

**Chair:**
**Professor Sir John Grimley Evans**,
Radcliffe Infirmary, Oxford

Ms Stella Burnside,
Altnagelvin Hospitals Trust,
Londonderry

Mr John Cairns,
University of Aberdeen

Professor Howard Cuckle,
University of Leeds

Dr Carol Dezateux,
Institute of Child Health,
London

Dr Anne Dixon Brown,
NHS Executive,
Anglia & Oxford

Professor Dian Donnai,
St Mary's Hospital, Manchester

Dr Tom Fahey,
University of Bristol

Mrs Gillian Fletcher,
National Childbirth Trust

Dr JA Muir Gray,
Institute of Health Sciences,
Oxford

Professor Alexander Markham,
St James's University Hospital,
Leeds

Dr Ann McPherson,
General Practitioner, Oxford

Dr Susan Moss,
Institute of Cancer Research

Dr Sarah Stewart-Brown,
University of Oxford

**Past members**

Dr Sheila Adam,
Department of Health*

Professor George Freeman,
Charing Cross & Westminster
Medical School, London

Dr Mike Gill, Brent & Harrow
Health Authority

Dr Anne Ludbrook,
University of Aberdeen

Professor Theresa Marteau,
Guy's, King's & St Thomas's
School of Medicine
& Dentistry, London

Professor Catherine Peckham,
Institute of Child Health,
London

Dr Connie Smith,
Parkside NHS Trust,
London

Ms Polly Toynbee, Journalist

Professor Nick Wald,
University of London

Professor Ciaran Woodman,
Centre for Cancer
Epidemiology, Manchester

# Primary and Community Care Panel

**Current members**

**Chair:**
**Dr John Tripp**,
Royal Devon & Exeter
Healthcare NHS Trust

Mr Kevin Barton,
East London & City
Health Authority

Professor John Bond,
University of Newcastle-
upon-Tyne

Dr John Brazier,
University of Sheffield

Ms Judith Brodie,
Age Concern, London

Mr Shaun Brogan,
Daventry & South Northants
Primary Care Alliance

Mr Joe Corkill,
National Association for
Patient Participation

Dr Nicky Cullum,
University of York

Professor Pam Enderby,
University of Sheffield

Mr Andrew Farmer,
Institute of Health Sciences,
Oxford

Professor Richard Hobbs,
University of Birmingham

Professor Allen Hutchinson,
University of Sheffield

Dr Phillip Leech,
Department of Health

Dr Aidan Macfarlane,
Oxfordshire Health Authority

Professor David Mant,
Institute of Health Sciences,
Oxford

Dr Chris McCall,
General Practitioner, Dorset

Dr Robert Peveler,
University of Southampton

Professor Jennie Popay,
University of Salford

Ms Hilary Scott,
Tower Hamlets Healthcare
NHS Trust, London

Dr Ken Stein,
North & East Devon
Health Authority

**Past members**

Professor Angela Coulter,
King's Fund, London*

Professor Martin Roland,
University of Manchester*

Dr Simon Allison,
University of Nottingham

Professor Shah Ebrahim,
Royal Free Hospital, London

Ms Cathy Gritzner,
King's Fund, London

Professor Andrew Haines,
RDRD, North Thames
Regional Health Authority

Dr Nicholas Hicks,
Oxfordshire Health Authority

Mr Edward Jones,
Rochdale FHSA

Professor Roger Jones,
Guy's, King's & St Thomas's
School of Medicine
& Dentistry, London

Mr Lionel Joyce,
Chief Executive, Newcastle City
Health NHS Trust

Professor Martin Knapp,
London School of Economics
& Political Science

Professor Karen Luker,
University of Liverpool

Dr Fiona Moss,
Thames Postgraduate Medical
& Dental Education

Professor Dianne Newham,
King's College London

Professor Gillian Parker,
University of Leicester

Dr Mary Renfrew,
University of Oxford

* Previous Chair

*continued*

# National Coordinating Centre for Health Technology Assessment, Advisory Group

**Current members**

**Chair:**
**Professor John Gabbay**,
Wessex Institute for Health Research & Development

Professor Mike Drummond,
Centre for Health Economics, University of York

Ms Lynn Kerridge,
Wessex Institute for Health Research & Development

Dr Ruairidh Milne,
Wessex Institute for Health Research & Development

Ms Kay Pattison,
Research & Development Directorate, NHS Executive

Professor James Raftery,
Health Economics Unit, University of Birmingham

Professor Ian Russell,
Department of Health Sciences & Clinical Evaluation, University of York

Dr Ken Stein,
North & East Devon Health Authority

Professor Andrew Stevens,
Department of Public Health & Epidemiology, University of Birmingham

**Past member**

Dr Paul Roderick,
Wessex Institute for Health Research & Development

# HTA Commissioning Board

## Current members

**Chair:**
**Professor Charles Florey**,
Department of Epidemiology & Public Health, Ninewells Hospital & Medical School, University of Dundee

Professor Doug Altman,
Director of ICRF/NHS Centre for Statistics in Medicine, Oxford

Professor John Bond,
Professor of Health Services Research, University of Newcastle-upon-Tyne

Mr Peter Bower,
Independent Health Advisor, Newcastle-upon-Tyne

Ms Christine Clark,
Honorary Research Pharmacist, Hope Hospital, Salford

Professor Shah Ebrahim,
Professor of Epidemiology of Ageing, University of Bristol

Professor Martin Eccles,
Professor of Clinical Effectiveness, University of Newcastle-upon-Tyne

Dr Mike Gill,
Director of Public Health & Health Policy, Brent & Harrow Health Authority

Dr Alastair Gray,
Director, Health Economics Research Centre, University of Oxford

Professor Mark Haggard,
MRC Institute of Hearing Research

Dr Jenny Hewison,
Senior Lecturer, Department of Psychology, University of Leeds

Professor Sir Miles Irving
(Programme Director),
Professor of Surgery, University of Manchester, Hope Hospital, Salford

Professor Alison Kitson,
Director, Royal College of Nursing Institute

Dr Donna Lamping,
Senior Lecturer, Department of Public Health, London School of Hygiene & Tropical Medicine

Professor Alan Maynard,
Professor of Economics, University of York

Professor Jon Nicholl,
Director, Medical Care Research Unit, University of Sheffield

Professor Gillian Parker,
Nuffield Professor of Community Care, University of Leicester

Dr Tim Peters,
Reader in Medical Statistics, Department of Social Medicine, University of Bristol

Professor Martin Severs,
Professor in Elderly Health Care, Portsmouth University

Dr Sarah Stewart-Brown,
Director, Institute of Health Sciences, University of Oxford

Professor Ala Szczepura,
Director, Centre for Health Services Studies, University of Warwick

Dr Gillian Vivian,
Consultant, Royal Cornwall Hospitals Trust

Professor Graham Watt,
Department of General Practice, Woodside Health Centre, Glasgow

Professor Kent Woods,
Regional Director of R&D NHS Executive, Trent

Dr Jeremy Wyatt,
Senior Fellow, Health & Public Policy, School of Public Policy, University College, London

## Past members

Professor Ian Russell,
Department of Health Sciences & Clinical Evaluation, University of York*

Professor David Cohen,
Professor of Health Economics, University of Glamorgan

Mr Barrie Dowdeswell,
Chief Executive, Royal Victoria Infirmary, Newcastle-upon-Tyne

Dr Michael Horlington,
Head of Corporate Licensing, Smith & Nephew Group Research Centre

Professor Martin Knapp,
Director, Personal Social Services Research Unit, London School of Economics & Political Science

Professor Theresa Marteau,
Director, Psychology & Genetics Research Group, Guy's, King's & St Thomas's School of Medicine & Dentistry, London

Professor Sally McIntyre,
MRC Medical Sociology Unit, Glasgow

Professor David Sackett,
Centre for Evidence Based Medicine, Oxford

Dr David Spiegelhalter,
MRC Biostatistics Unit, Institute of Public Health, Cambridge

Professor David Williams,
Department of Clinical Engineering, University of Liverpool

Dr Mark Williams,
Public Health Physician, Bristol

* Previous Chair

*Health Technology Assessment* 1999; Vol. 3: No. 6

Copies of this report can be obtained from:

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 1703 595 639 Email: hta@soton.ac.uk
http://www.soton.ac.uk/~hta

ISSN 1366-5278