OXFORD

HANDBOOKS IN HEALTH ECONOMIC EVALUATION

ECONOMIC EVALUATION IN CLINICAL TRIALS

SECOND EDITION

Henry A. Glick | Jalpa A. Doshi | Seema S. Sonnad | Daniel Polsky

Economic Evaluation in Clinical Trials

Handbooks in Health Economic Evaluation Series

Series editors: Alastair Gray and Andrew Briggs

Existing volumes in the series:

Decision Modelling for Health Economic Evaluation
Andrew Briggs, Mark Sculpher, and Karl Claxton

Applied Methods of Cost-Effectiveness Analysis in Healthcare
Alastair M. Gray, Philip M. Clarke, Jane L. Wolstenholme, and Sarah Wordsworth

Applied Methods of Cost–Benefit Analysis in Healthcare
Emma McIntosh, Philip Clarke, Emma Frew, and Jordan Louviere

Economic Evaluation in Clinical Trials

SECOND EDITION

Dr. Henry A. Glick

Professor of Medicine, Division of General Internal Medicine, Perelman School of Medicine; Professor of Health Care Systems, The Wharton School; Senior Fellow, Leonard Davis Institute of Health Economics; Associate Scholar, Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, United States; Visiting Professor, Institute of Health and Wellbeing, University of Glasgow, United Kingdom

Dr. Jalpa A. Doshi

Associate Professor of Medicine, Division of General Internal Medicine, Perelman School of Medicine; Director, Economic Evaluations Unit, Center for Evidence-Based Practice; Director, Value-Based Insurance Design Initiatives; Center for Health Incentives and Behavioral Economics; Senior Fellow, Leonard Davis Institute of Health Economics, University of Pennsylvania, United States

Dr. Seema S. Sonnad

Director of Health Services Research, The Value Institute, Christiana Care Health System, United States; Adjunct Associate Professor, Department of Surgery, Perelman School of Medicine, Adjunct Senior Fellow, Leonard Davis Institute of Health Economics, University of Pennsylvania, United States

Dr. Daniel Polsky

Robert D. Eilers Professor, Health Care Management, The Wharton School; Professor of Medicine, Division of General Internal Medicine, Perelman School of Medicine; Executive Director, Leonard Davis Institute of Health Economics, University of Pennsylvania, United States

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2007
Second Edition published in 2015

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2014940458

ISBN 978–0–19–968502–8

Printed in Great Britain by
Clays Ltd, St Ives plc

Series preface

Economic evaluation in healthcare is a thriving international activity that is increasingly used to allocate scarce health resources, and within which applied and methodological research, teaching, and publication are flourishing. Several widely respected texts are already well established in the market, so what is the rationale for not just one more book, but for a series? We believe that the books in the series Handbooks in Health Economic Evaluation share a strong distinguishing feature, which is to cover as much as possible of this broad field with a much stronger practical flavor than existing texts, using plenty of illustrative material and worked examples. We hope that readers will use this series not only for authoritative views on the current practice of economic evaluation and likely future developments, but for practical and detailed guidance on how to undertake an analysis. The books in the series are textbooks, but first and foremost they are handbooks.

Our conviction that there is a place for the series has been nurtured by the continuing success of two short courses we helped develop—Advanced Methods of Cost-Effectiveness Analysis, and Advanced Modelling Methods for Economic Evaluation. Advanced Methods was developed in Oxford in 1999 and has run several times a year ever since, in Oxford, Canberra, and Hong Kong. Advanced Modelling was developed in York and Oxford in 2002 and has also run several times a year ever since, in Oxford, York, Glasgow, and Toronto. Both courses were explicitly designed to provide computer-based teaching that would take participants through the theory but also the methods and practical steps required to undertake a robust economic evaluation or construct a decision-analytic model to current standards. The proof-of-concept was the strong international demand for the courses—from academic researchers, government agencies and the pharmaceutical industry—and the very positive feedback on their practical orientation.

So the original concept of the Handbooks series, as well as many of the specific ideas and illustrative material, can be traced to these courses. The Advanced Modelling course is in the phenotype of the first book in the series, *Decision Modelling for Health Economic Evaluation,* which focuses on the role and methods of decision analysis in economic evaluation. The Advanced Methods course has been an equally important influence on *Applied Methods of Cost-Effectiveness,* the third book in the series which sets out the key elements

of analyzing costs and outcomes, calculating cost-effectiveness, and reporting results. The concept was then extended to cover several other important topic areas. First, the design, conduct, and analysis of economic evaluations alongside clinical trials have become a specialized area of activity with distinctive methodological and practical issues, and its own debates and controversies. It seemed worthy of a dedicated volume, hence the second book in the series, *Economic Evaluation in Clinical Trials.* Next, while the use of cost–benefit analysis in healthcare has spawned a substantial literature, this is mostly theoretical, polemical, or focused on specific issues such as willingness to pay. We believe the fourth book in the series, *Applied Methods of Cost-Benefit Analysis in Health Care*, fills an important gap in the literature by providing a comprehensive guide to the theory but also the practical conduct of cost–benefit analysis, again with copious illustrative material and worked out examples.

Each book in the series is an integrated text prepared by several contributing authors, widely drawn from academic centers in the United Kingdom, the United States, Australia, and elsewhere. Part of our role as editors has been to foster a consistent style, but not to try to impose any particular line: that would have been unwelcome and also unwise amidst the diversity of an evolving field. News and information about the series, as well as supplementary material for each book, can be found at the series website: <http://www.herc.ox.ac.uk/books>.

Alastair Gray
Oxford

Andrew Briggs
Glasgow

Web resources

In addition to worked examples in the text, readers of this book can download datasets and programs for Stata® for Windows (Stata Corporation, College Station, Texas, United States) that provide examples of the analysis of cost and quality-adjusted life years, estimation of sampling uncertainty for the comparison of cost and effect, and calculation of sample size and power for cost-effectiveness analysis in clinical trials.

Materials for the book are maintained at the following web addresses: <http://www.uphs.upenn.edu/dgimhsr/eeinct.htm and <http://www.herc.ox.ac.uk/books/trials.shtml>.

More information is available on the websites. We anticipate that the web-based material will be expanded and updated over time.

Acknowledgment

I first became involved in economic assessment in clinical trials in 1985 when John M. Eisenberg (1946–2002) hired me to manage an economic evaluation for a Veterans Administration (VA) cooperative trial of total parenteral nutrition. I began thinking about the ideas in this book the next year when two events occurred. First, John and I began co-teaching a graduate course in the School of Medicine that eventually was named Medical Decision Making and Clinical Economics. Since 1990, I have co-taught this course with Sankey Williams. Second, Mark Pauly and I began co-teaching Cost-Benefit and Cost-Effectiveness Analysis, a graduate course in the Wharton School. Except for Mark's sabbatical here and there, Mark and I continued to co-teach it, eventually with Dan Polsky, until 2012. Dan, Ashley Swanson and I now teach the course together. John, Mark, Sankey, and Dan have all challenged me and helped me grow both as a researcher and as a person.

I can trace some of the ideas in this book to early lecture notes from these classes. The hundreds of students who have listened to me, questioned me, and made me rethink ways to explain many of the ideas that we present here have all made this a better book than it otherwise would have been.

Other ideas, particularly those in Chapters 5, 8, and 9, were developed as part of research that was supported by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) grant 1 R0l AA12664–01A2 and the National Institute of Drug Abuse (NIDA) grant R0l DA017221–0lAl.

Early in my career, Eugene Smith, Joe Heyse, John Cook, Dick Willke, and Martin Backhouse all provided me with funding and opportunities to work on trials. So too did the International Clinical Epidemiology Network (INCLEN). Kevin Schulman was my earliest collaborator, and I think of him as one still. Dan was and is my second collaborator and Jalpa Doshi is the third. Bruce Kinosian and I have worked together since shortly after I began working in the Department of Medicine; Andy Briggs has always had more confidence in me than I deserve. They have all helped me gain whatever success I have had in my career.

My Mom who died in 1984 and my Dad who died in 2008 always had faith in me. I hope they would have been proud. Even though I do not see him, Avivar, my son, brings a smile to my face every day. Finally, there is Seema who helps me through each day. Thank you all.

Henry A. Glick

Contents

Chapter 1

Introduction to economic evaluations in clinical trials

1.1 Introduction

During the past 30 years, healthcare expenditures have increased dramatically throughout the world. In the US, total national expenditures on healthcare as a percentage of gross domestic product (GDP) have increased from 7.4% (equivalent to $436 per capita) in 1972 to 17.7% (equivalent to $8564 per capita) in 2011 [1, 2]. Dramatic increases have also taken place in many other developed countries. During this same time period, total expenditures as a percentage of GDP have increased in the UK from 4.6% to 9.4%, in France from 6.2% to 11.6%, in Germany from 7.1% to 11.3%, and in Japan from 4.8% to 10.0% [1, 2].

In response to these increases, countries around the world have been investigating methods to control healthcare cost. Macro-level methods have included greater risk sharing between payers, providers, and patients as well as a greater reliance on market-oriented incentives. Micro-level methods have included making decisions about the utilization of particular medical therapies by government regulators, healthcare providers, members of formulary committees, payers, and patients based on an evaluation of value for the cost of those therapies.

One potential source of information for decisions about value for the cost derives from clinical trials that establish the efficacy and effectiveness of medical therapies. During the past 25 years, there has been a growing trend to exploit this potential source of information by collecting data on medical service use, cost, and effect in clinical trials. Most frequently economic evaluation has been incorporated into the drug development process, for example, in phase III and sometimes phase II, during which a drug's safety and efficacy are evaluated prior to regulatory approval, as well as in phase IV, which occurs after the drug is marketed. Economic evaluations are also increasingly being conducted within trials of other medical therapies such as surgical procedures, behavioral interventions, etc. The UK Medical Research Council and US National Institutes of Health routinely request the inclusion of economic assessments prior to

funding large-scale multicenter trials. Many other countries require evidence of economic value as part of their reimbursement decision-making, and clinical trials provide one of the earliest opportunities to generate economic data that can be used for this purpose.

Coincident with this increased attention has been a rapid development in the methodologies used to evaluate the value for the cost of new therapies. More than 25 years ago, most economic evaluations derived from clinical trials were not based on direct observation of the impact of a therapy on cost and effect. Rather, they were based on decision analyses that were developed primarily from epidemiological data, and borrowed only a few key findings, such as the odds ratio for effectiveness, from a clinical trial. Reported results included point estimates of incremental cost and effect and a point estimate for the cost-effectiveness ratio. Uncertainty was expressed solely by the use of sensitivity analysis.

By the mid-1990s, a growing number of trial-based economic evaluations were based on direct observation of the impact of a therapy on cost and effect. In their 1998 review, Barber and Thompson reported that they identified 45 cost-effectiveness analyses published in 1995 that were based on the analysis of patient-level data that were collected as part of a randomized clinical trial [3]. In these studies, short-term economic impacts of the therapy were directly observed; longer-term impacts could potentially be projected by the use of a range of extrapolation methods including decision analytic models.

As with the earlier economic studies, these studies continued to report point estimates of incremental cost and effect, but they also had the opportunity to report confidence intervals for these differences, based on univariate tests of means or multivariable ordinary least squares (OLS) regression. Most, if not all, of the assessments of value for the cost would have continued to be based on point estimates and sensitivity analysis, in part because the first major article that reported on methods for addressing sampling uncertainty related to cost-effectiveness ratios was published in 1994 [4]. Finally, little to no consideration would have been given to the transferability of the economic results of the trials.

In the past 15 years, the field has continued to mature. Doshi et al. [5] reported that in 2003, the number of published economic evaluations that were based on the analysis of patient-level data on cost and effect collected as part of a randomized clinical trial had increased to 115. Sculpher et al. [6] have found that "Since 1994, approximately 30% of published economic evaluations on the NHS Economic Evaluation Database have been based on data from a single RCT" (p. 677). A number of studies are now reporting point estimates and confidence intervals for incremental cost, incremental effect, and for the evaluation of value for the cost, for example, by reporting the confidence interval for

the cost-effectiveness ratio or for net monetary benefit. By this time, the impact of sensitivity analysis on the comparison of cost and effect could be judged by how it affected both the point estimate of the cost-effectiveness ratio or net monetary benefit as well as its effects on the resulting confidence intervals.

1.2 Steps in conducting an economic evaluation in clinical trials

The steps involved in conducting an economic evaluation that is incorporated within a randomized trial include: (1) quantification of the cost and effect of care; (2) assessment of whether and by how much average cost and effect differ among the treatment groups; (3) comparison of the magnitudes of differences in cost and effect and evaluation of the "value for the cost" of the therapies, for example, by reporting incremental cost-effectiveness ratios; and (4) identifying the populations to whom the results apply. We follow this same structure in the remainder of this volume.

As in the first edition, in Chapter 2, we discuss issues related to the design of economic evaluations in trials. We address six study design issues, which include: (1) what preplanning should be done in preparation for the trial? (2) What medical service use should we measure? (3) At what level should service use be aggregated? (4) Which price weight estimates should be used for the study? (5) How naturalistic should the study design be? (6) What should we do if the full benefit and cost of therapy are not expected to be observed during the trial? For this and the other chapters, we have updated our literature review as well as the data that we report. In addition, for Chapter 2 we have added or updated material about instruments that have been used to collect information on the use of medical services, the accuracy of patient recall of these services, instruments that have been used to collect information on work loss (both absenteeism and presenteeism), and protocol-induced costs. Making sound design decisions about the length of economic follow-up, sample size, minimization of lost-to-follow-up of economic data, collection of adequate amounts of data on medical service use, and the like are essential for trials to provide useful information at their conclusion.

Once data on medical service use are collected, they commonly are translated into measures of cost by multiplying service use by price weights. In Chapter 3, we discuss issues related to the selection of an appropriate set of these weights. Studies commonly use either national or center-specific price weights. Center-specific weights are more likely to provide an estimate of the cost that was incurred in the trial. National price weights are generally considered to be more representative than center-specific weights, particularly for questions of national resource allocation. However, national representativeness of the

economic results depends on representative patterns of both medical service use and price weights. In addition to general updating, in this chapter we have added a more complete discussion of sources of medication costs, attempted to simplify the discussion of the impacts of substitution effects in Box 3.2, and expanded the discussion of inflation indices.

Potential measures of economic effect range from intermediate outcomes, such as blood pressure in millimeters of mercury, to final outcomes, such as length of survival. The Panel on Cost Effectiveness in Health and Medicine has recommended the use of quality-adjusted life years (QALYs) as the principal measure of effect in cost-effectiveness analysis [7, p. 308]. QALYs have the advantages that they combine multiple dimensions of outcome, that is, survival and quality of life, into a single measure that allows comparisons to be made across therapeutic areas and illnesses and for which we generally understand how much we are willing to pay for a unit of effect. In Chapter 4, we discuss methods for assessing QALYs. In particular we focus on the two most common general approaches to QALY assessment, the use of prescored instruments and direct elicitation methods. Although our conclusions remain relatively unchanged, in this edition we have made a major update of our comparison of these different instruments (including the addition of the prescored SF-6D). We have also attempted to simplify our discussion about the role of responsiveness in instrument selection and added sections on minimally important differences for these instruments, methods of administration of the instruments, and disease-specific, preference-weighted quality of life instruments.

In Chapters 5 and 6, we explore methods for the analysis of cost. In Chapter 5, we discuss strategies for this analysis when we are not concerned about large amounts of censored data; in Chapter 6, we discuss strategies for addressing censored data. The central themes of these chapters are that: (1) the outcome of interest for economic assessment is the difference in the sample mean (arithmetic mean) of cost (and the difference in sample mean of effect), (2) the analytic methods we adopt should estimate and yield inferences about this difference and should not estimate and yield inferences about differences in geometric means or medians, and (3) the evaluation of the difference in the sample mean of cost is complicated by the fact that the distribution of cost is typically highly skewed with long heavy right tails. As we indicate in Chapter 6, censoring poses additional problems that require diagnosis of the mechanism of censoring and adoption of methods that are appropriate for analysis given the diagnosed mechanism. Updates in Chapter 5 include a discussion of the problems with interpreting the coefficients of log OLS as representing percentage differences, expansion of the description of bootstrapping, a more detailed description of recycled predictions, and the addition of a brief section on the

analysis of preference scores. Updates in Chapter 6 focus primarily on an update of the literature.

Once we have analyzed cost and effect, we compare the two. In Chapter 7, we introduce the two principal methods available for this comparison: the cost-effectiveness ratio and net monetary benefit (NMB). In Chapters 8–10, we address two of the potential limitations in the interpretation of the comparison of cost and effect observed in clinical trials. In Chapters 8 and 9, we discuss the concepts underlying the measurement of sampling uncertainty for cost-effectiveness ratios and NMB (Chapter 8) as well as how to calculate these measures of uncertainty (Chapter 9). Discussion of the latter issue also provides us with the opportunity to address issues of sample size and power for economic evaluations that are incorporated into clinical trials. In Chapter 10, we address the issue of transferability: to whom does the pooled result from the trial apply?

Updates to Chapters 7–9 include discussion of a selection algorithm for NMB (Chapter 7) and calculation and interpretation of the value of information (Chapters 8 and 9); a major revision (and, we hope, clarification) of the discussion of experiment 2 (Chapter 8); addition of greater detail about ordering bootstrap replicates when constructing nonparametric confidence intervals for the cost-effectiveness ratio (Chapter 9); discussion of the problems that can arise when constructing multi-therapy acceptability curves (Chapter 9); and description of the role played by negative values of willingness to pay in the interpretation of NMB graphs, acceptability curves, value of information curves, and power curves (Chapters 8 and 9). In addition to updating references, the primary addition to Chapter 10 is a listing of checklists that have been proposed for assessing transferability of the economic results of trials.

As we indicated above, economic data collected as part of clinical trials are one potential source of information for decisions about value for the cost, but they are not the only potential source of these data. Data from trials and from decision analyses are complementary and have different strengths and weaknesses. We are more certain about what we observed during the trial, but limitations in terms of length of follow-up, treatment comparators, or patient populations studied may mean that the trial does not address all of the considerations that go into treatment adoption decisions. We are less certain about the results from decision analysis, in part, because they either cannot be or have not been formally validated or empirically tested. But these models allow us to address the broader set of considerations that may go into adoption decisions.

There are other volumes in this series that address decision analysis, and they are appropriate for efforts you may make in the development of these models. This volume is intended to provide you with an in-depth understanding of economic assessment conducted by the use of patient-level data collected as part of

randomized controlled trials. Our goal is to describe methods that maximize the information we can derive from such patient-level analysis. In doing so, we highlight good practice, we identify commonly misused practices, and in those cases where methods are still rapidly evolving, we summarize the alternative methods.

References

1 Organisation for Economic Cooperation and Development. *OECD Health Data 98*. Paris: OECD Electronic Publications, 1998.

2 The World Bank. *World Development Indicators. Health Expenditure, Total (% of GDP)*. Available at: <http://data.worldbank.org/indicator/SH.XPD.TOTL.ZS>.

3 **Barber JA, Thompson SG.** Analysis and interpretation of cost data in randomised controlled trials: review of published studies. *BMJ*. 1998; **317**: 1195–200.

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

5 **Doshi JA, Glick HA, Polsky DP.** Analyses of cost data in economic evaluations conducted alongside randomized controlled trials. *Value in Health*. 2006; **9**: 334–40.

6 **Sculpher MJ, Claxton K, Drummond M, McCabe C.** Whither trial-based economic evaluation for health care decision making? *Health Econ*. 2006; **15**: 677–87.

7 **Gold MR, Siegel JE, Russell LB, Weinstein MC (eds).** *Cost-effectiveness in Health and Medicine*. Oxford: Oxford University Press, 1996.

Chapter 2

Designing economic evaluations in clinical trials

2.1 Introduction to designing economic evaluations in clinical trials

This chapter provides an introduction to issues of design of economic evaluations conducted as part of randomized trials. We first describe the gold standard economic evaluation and the tensions that exist in the design of such a study. Second, we discuss six study design issues related to these evaluations. Finally, we address issues such as when it is inappropriate to perform an economic evaluation within a trial and features that contribute to a successful economic evaluation within a trial.

2.2 The gold standard

The gold standard economic evaluation within a clinical trial has a number of defining characteristics. First, it is conducted in naturalistic settings, uses as a comparator a commonly used, cost-effective therapy, and studies the therapy as it would be used in usual care. Second, it is performed with adequate power to assess the homogeneity of economic results in a wide range of clinical settings and among a wide range of clinical indications in which the therapy will be used. Third, it is designed with an adequate length of follow-up to assess the full impact of the therapy. Fourth, it is conducted within a time frame that allows the resulting data to inform important decisions in the adoption and dissemination of the therapy.

In a gold standard evaluation, we measure all cost of all participants in the trial regardless of why the cost was incurred, starting prior to randomization and continuing for the duration of follow-up of the trial. Cost incurred after randomization constitutes the cost outcome of interest in the trial (at least for the period of observation within the trial; in some cases, we may also want to project cost and effect beyond this period of observation). Cost incurred during a prespecified period prior to randomization (e.g., 1, 6, or 12 months) is a potential predictor of cost after randomization, and will likely explain variability in this

cost. The optimal length of this prior period is generally disease-dependent and should be based on the collection of sufficient data to differentiate high utilizers from low. While it is best to collect the same types of data as will be collected during follow-up, burden can be reduced by identifying a subset of these services for data collection.

Given that in the gold standard we measure all cost, we measure the cost that is believed to be related to the disease and its treatment as well as the cost that might be expected to be unrelated. We do so because the gold standard evaluation is adequately powered; thus, rather than making potentially flawed judgments about the relationship between therapy and a particular medical service, we empirically determine the incremental cost that is related to the therapy. This measurement of all cost has implications for the debate about whether or not the evaluation should include only that cost that is related to the intervention or whether it should include "unrelated" cost as well [1, 2].

Issues related to the type of analysis that will be conducted, for example, cost–benefit, cost-effectiveness, or cost-minimization analysis, apply equally to economic evaluations that are incorporated within clinical trials and to other types of economic evaluation. So too do issues related to the types of cost that will be included, for example, direct medical, direct nonmedical, productivity, and intangible, and the perspective from which the study will be conducted. These issues have been well addressed in the literature [3–5], and we do not discuss them further.

Performing a gold standard evaluation is most feasible when it is easy to identify when services are provided, for example, in hospital-based studies or studies conducted within integrated systems of care. Feasibility is also enhanced when medical service use and cost data are already being collected, for example, in an administrative database, and when researchers have ready access to these data once they are collected. Traditionally, hospital-based studies have been considered as some of the best candidates for gold standard studies, because of the closed nature of the hospital, and because of the large amounts of record-keeping that are carried out there. However, as episodes of care increasingly continue after discharge from the hospital, and when some types of medical service use are not collected as part of the administrative record, for example, physician cost in the US, even these studies fail to meet the standard.

At the same time, the requirements of gold standard assessments pose several drawbacks and feasibility issues. First, there may be contradictions between the goal related to conducting the trial in naturalistic settings and the one related to providing information within a time frame that informs important decisions. The information from trials that are performed early in the life of a therapy, for example, during phases II and III of the drug development process, may aid

decisions about early adoption and diffusion of the therapy. If the therapy is dramatically effective, these trials may also be one of the last chances we have to randomize patients to receive the therapy, because once information about a therapy's clinical effectiveness is available, patients may not be willing to participate in experiments simply to evaluate its value for the cost. These early trials, however, may not reflect the cost and effect that would be observed in usual practice, in part because, due to regulatory needs, the trials may affect usual practice and in part because clinicians may not yet know how to use the therapy efficiently.

Trials that are performed later, such as post-marketing studies in the drug development process, are more capable of reflecting usual practice, although it is possible that—due to formal protocols—such trials yield biased results and/or reduce observed variation in practice. However, information from these trials may be too late to inform important early decisions about the adoption and diffusion of the therapy, and as noted, if the therapy is shown to be clinically effective, patients may be unwilling to enroll in them. To address some, but not all, of these issues, ideally, we would evaluate new therapies throughout the development process, to inform early decisions and to re-evaluate economic findings once the therapy is in common use.

A second feasibility issue is that the need for adequate power in gold standard economic evaluations may require a larger number of study participants than the investigators or funders are willing to enroll. Third, such evaluations may require a longer follow-up than that required for clinical endpoints; for example, the economic impacts of therapy may not end at 28 days, but regulatory agencies may accept clinical endpoints measured at 28 days as evidence of clinical efficacy. Finally, the additional burden of economic data collection may exacerbate the problem that some investigators already face in collecting the clinical data required for a trial. Reconstructing the equivalent of a patient bill in the case report form is burdensome. Also, investigators may have limited access to information on medical services provided to study participants from providers and in centers that are unaffiliated with the trial.

Because of these limitations, investigators often make trade-offs between the ideal economic assessment and assessments that are most feasible under the presumption that imperfect information is better than none.

2.3 **Six study design issues**

When designing a study, we need to consider at least six sets of issues. These include: (1) what preplanning should be done in preparation for the trial? (2) What medical service use—also referred to as resource use—should we

measure? (3) At what level should service use be aggregated? (4) Which price weight—also referred to as unit cost—estimates should be used for the study? (5) How naturalistic should the study design be? (6) What should we do if the full benefit and cost of therapy will not be observed during the trial?

2.3.1 What preplanning should be done in preparation for the trial?

A number of preplanning activities should be performed when designing an economic assessment within a clinical trial. These include, but are not limited to, identifying an appropriate length of follow-up for economic endpoints; estimating arithmetic means, variances, and correlations for cost, health-related quality of life, and preference; identifying the types of medical services used by study participants; identifying and evaluating data collection instruments and procedures; and gauging levels of patient interest in the study.

Identifying an appropriate length of follow-up for economic endpoints

Economic assessments conducted as part of randomized trials are meant to help decision-makers reach conclusions about the economic benefits of the therapy under investigation. One design issue that may limit the interpretability of the economic data collected within the trial is the study's time horizon. Although clinical efficacy may be demonstrated when a difference in clinical endpoints is observed between study arms, from an economic perspective the appropriate time horizon for a trial would include all or a substantial portion of the time when there is medical service use related to the illness under study. The economic time horizon that would best inform decision-makers about the value for the cost of a therapy thus need not be the same as the one adopted for answering the clinical question.

A number of approaches are available for identifying an appropriate economic time horizon for a clinical trial. One approach is to identify the economic episode of care. As defined in the literature, an episode of care is the period initiated by patient presentation with a diagnosis of a clinical condition, or in the case of a randomized trial, initiated by randomization, and concluded when the condition is resolved [6–9]. For definition of the clinical episode, resolution may refer to the acute clinical condition; for definition of the economic episode, it may refer to the return of cost or other outcomes such as preference for current health to the level that would have existed had the clinical condition not been present [10].

Mehta et al. [11] and Schulman et al. [12] have used episode of care methodology to define economic episodes of care for diabetic foot ulcers [11] and

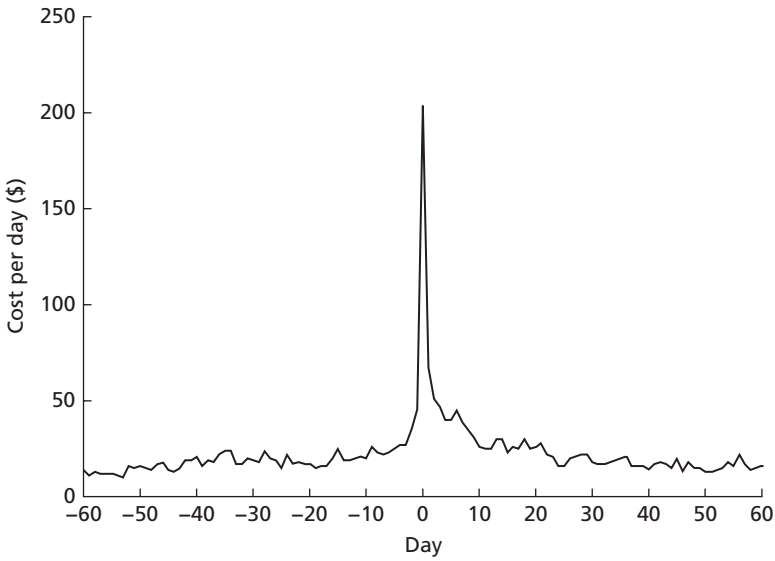

Fig. 2.1 Migraine cost [13].

Reproduced from Glick HA, Polsky D, and Schulman K, Trial-based Economic Evaluations: An Overview of Design and Analysis, Figure 6.2, p.119, in Drummond M and McGuire A (eds), *Economic Evaluation in Health Care: Merging Theory and Practice*, Oxford University Press, Oxford, UK, Copyright © 2001, by permission of Oxford University Press. Source: data from Pennsylvania Medicaid.

migraine headaches [12]. These authors quantified the length of an economic episode by comparing differences in mean daily cost and differences in the proportion of patients with cost before and after an index diagnosis of foot ulcer and migraine headache: the episode was said to last until the initially elevated mean cost per day (proportion of patients with cost) after the diagnosis returned to their level prior to the diagnosis.

Fig. 2.1 shows the mean daily cost for 60 days prior to an index diagnosis of migraine (days −60 to −1), the mean cost during the day of the index diagnosis (day 0), as well as mean daily cost for the 60 days after the index diagnosis (days 1–60). Analysis of the data suggested that cost from a migraine headache may remain elevated for 3 weeks [12], which is markedly longer than the length of clinical follow-up for migraine therapies, which often can be as short as 30–120 min. Given that the economic evaluation is meant to identify both the clinical and economic impacts of therapies, in the case of treatment for migraine, we would be more likely to capture the latter effects by lengthening the time horizon of trial follow-up and studying the entire period of time when cost is elevated.

Decision analysis represents a second approach that can be adopted for identifying an appropriate economic time horizon for a clinical trial [14]. For example, a decision analytic model that simulates the disease and the effects of therapy might demonstrate that life expectancy among individuals who would have died without the intervention but now live because of it has a substantial impact on the economics of the therapy. In this case, the time horizon of the trial could be adjusted so that we are able to assess these gains in life expectancy, or we could ensure that data are collected that would allow the prediction of life expectancy after the trial.

Estimating arithmetic means, variances, and correlations

Arithmetic means, variances, and correlations for cost, health-related quality of life, and preference provide the information necessary for assessing the sample size required to answer the economic questions posed in the study. Enrollment size for randomized trials often is based on the minimum number of study participants needed to address the clinical questions. However, the number of participants required for the economic assessment may differ from that needed for the clinical evaluation.

Prior to the development of the literature that described confidence intervals for cost-effectiveness ratios [15–19], a common approach to sample size calculation for economic evaluation in trials was to select the larger of the sample sizes needed for estimating pre-specified cost and effect differences. For example, what sample size was required to identify a significant difference in cost of 1000, and what was required to identify a significant 10% reduction in mortality?

The development of this literature made it clear, however, that the goal of economic evaluation in trials was to determine the level of confidence we can have about a therapy's value for the cost. Thus, while infrequently used [20], current sample size methods base their calculations on the number of study participants needed to: (1) rule out unacceptably high cost-effectiveness ratios [21], (2) rule out that the net monetary benefits of the intervention are less than 0, or (3) maximize the value of information from the trial [22]. See Chapter 9 for formulas for sample size and power.

These methods generally require more information than is needed for estimating sample sizes for clinical outcomes or for cost differences alone. Basic data for such calculations include the magnitude of the incremental cost and effect we expect to observe in the trial; the standard deviations for cost and effect in each of the treatment groups; our willingness to pay for health, for example, 30,000 or 50,000 per quality-adjusted life year (QALY); and the correlation between the difference in cost and effect. Although the amount of data

for the two types of sample size calculation may differ, the principles we use to generate them are the same. We may have access to published or unpublished data. In cases where such data are unavailable, we usually resort to making plausible assumptions and performing sensitivity analysis. While it is true that some investigators may start with a convenient/feasible sample size and generate potentially plausible assumptions to justify it, that is not the purpose of sample size calculation.

Where do we obtain this information about means, standard deviations, and correlations? If both therapies are already in use, expected differences in outcome and standard deviations can be derived from previous studies, new feasibility studies, or records of patients like those who will be enrolled in the trial. At least one study has suggested that the correlation between cost and effect observed in these data may be an adequate proxy for the correlation between the difference in cost and effect [15]. Alternatively, we might derive a guesstimate of the correlation by considering the expected relationships between intervention cost, other cost, and outcome. Is intervention cost likely to be positively associated with both other cost and outcome (e.g., if incurring greater intervention cost saves more lives and induces more care)? In that case, the correlation is likely to be positive. Is it instead negatively associated with other cost and outcome (e.g., if incurring greater intervention cost is due to treatment failure)? In that case, the correlation is likely to be negative.

For novel therapies that have yet to be used in large numbers of patients, information about the magnitude of the incremental cost and effect may not be available, and thus may need to be generated by assumption. Data on the standard deviations for those who receive usual care/placebo again may be obtained from feasibility studies or from patient records. We may assume that the standard deviation thus obtained will apply equally to both treatment groups, or we may make alternative assumptions about the relative magnitudes of these deviations. The observed correlation between cost and effect might again be used as a proxy for the correlation between the differences. (See Glick [21] for additional discussion.)

In some instances, the economic outcome may have less power than the clinical outcome. In this case, power calculations can be used to assess the magnitude of the cost and effect differences that can be detected. However, due to the fact that economic outcomes are joint outcomes of cost and effect, it is also possible for them to have more power than clinical outcomes [21, 23]. For example, it is possible for the p value for the difference in both cost and effect to be greater than 0.05, yet for us to confidently conclude that the ratio falls below 30,000 or 50,000 or, alternatively, that net monetary benefit—calculated by use of a willingness to pay of 30,000 or 50,000—is significantly greater than 0.

The discussion of sample size for addressing economic questions has implications for the selection of endpoints for the economic study. As indicated, determination of sample size requires that we be able to specify what we are willing to pay for a unit of health effect. Debate remains about willingness to pay for a QALY [24–28], which—except for the US Congress [29]—is the currently recommended outcome for cost-effectiveness analysis [5]. When researchers use disease-specific measures of effect such as cases of disease detected, or in alcohol research, abstinence days, or in many fields, symptom-free days, it is even less clear how much we should be willing to pay for these outcomes. While we can calculate a cost-effectiveness ratio for any outcome we want, to be convincing that a more costly and more effective therapy represents good value, the outcome must be one for which we have recognized benchmarks of cost-effectiveness. These considerations argue against use of too disease-specific an outcome for economic assessment, or at least for reporting both disease-specific and QALY outcomes.

As an example, Gupta [30] reported data indicating that in comparison with a topical antifungal agent, one triazole antifungal agent cost $5.67 per incremental toenail fungus symptom-free day. In comparison with this first triazole agent, the second triazole compound cost $8.88 per incremental symptom-free day. Without an idea of the value of a toenail fungus symptom-free day, it is not clear whether we should reject both of the triazole antifungal agents, for example, if our willingness to pay is less than or equal to $5.67; adopt the less expensive and less effective of the two triazole compounds if our willingness to pay is greater than $5.67 and less than or equal to $8.88; or adopt the higher cost, more effective triazole compound if our willingness to pay is greater than $8.88.

Identifying the types of medical services used by study participants

A third preplanning activity is to identify the types of medical services that are likely to be used by study participants. We can do so by reviewing medical charts or administrative data sets, having patients keep logs of their medical service use, or asking patients and experts about the kinds of care received by those with the condition under study. This information will help answer questions about the medical service use that should be measured (see discussion in section 2.3.2), particularly if, because of the burden of economic data collection, only a subset of medical services are to be recorded in case report forms. While analysis of administrative data may provide good estimates of the types of medical services that are used in current practice, we must also be prepared to collect data on additional services to account for the fact that new therapies may change practice, for example, by causing adverse outcomes that require different kinds of services than are used by patients in usual practice.

Identifying and evaluating data collection instruments and procedures

A number of instruments for collecting data on use of medical services have been published. Some include the Client Socio-Demographic and Service Receipt Inventory (CSDSRI) [31], the Resource Use in Dementia (RUD) instrument [32], the Resource Use Inventory (RUI) [33], and the questionnaire by Chernyak et al. [34]. There are also a large number of instruments that have been developed for particular studies. A number of these are catalogued in the Database of Instruments for Resource Use Measurement (<http://www.dirum.org>) [35].

As with any other forms and procedures used in clinical trials, those used to collect economic data should be pilot tested for their efficiency, clarity, and ease of use. Poorly designed forms can lead to low-quality data that will jeopardize our ability to draw useful conclusions from the study [36].

Gauging levels of patient interest in the study

Substantial amounts of data may be collected from study participants, for example, use of medical services from providers and in centers that are unaffiliated with the trial, patient out-of-pocket and transportation costs, quality of life data, and measures of patient preference. The amount that is feasible to collect will depend on the level of the participants' interest in the study. This level of interest potentially varies from disease to disease, and gauging it allows investigators to better estimate how much data collection the study participants will tolerate.

2.3.2 What medical service use should we measure?

In a trial that provides ready access to administrative data, for example, from insurers, government agencies, or hospitals, few if any trade-offs need to be made about the proportion of medical service use—and thus cost—that should be measured in the study. However, if data on this use will be collected prospectively in study case report forms, then the goals are to: (1) measure services that make up a large portion of the difference in treatment between study participants randomized to the therapies under evaluation—sometimes referred to as cost "drivers"—and (2) measure services that make up a large portion of the total cost of care. The former services provide an estimate of the new therapy's impact on the cost of care; the latter services provide a measure of overall variability of cost. In addition, failure to measure a large portion of the total cost of care leaves the study open to criticisms that differences in measured services observed between therapies may have been offset among services that were not measured in the study.

The best approach is to measure as many services as possible, because minimizing the services that go unmeasured reduces the likelihood that differences among them will lead to study artifacts. However, there are neither a priori guidelines about how much data are enough, nor data on the incremental value of specific items in the economic case report form. Decisions about those procedures that will be recorded within the case report form and those that will not should take into account the expense of collecting particular data items.

We can obtain information to guide these decisions if, during the preplanning process, we identify the types of medical services that are used by patients who are similar to the patients who will be enrolled in the trial. In general, however, substantially more experience with various medical problems and interventions is still needed before we know which data are essential to document in an economic case report form.

The root of the concern about the amount and type of medical service use that should be collected within a trial is that there may be insufficient resources to identify all medical service use among all participants. In contrast to our general advice, several alternative strategies have been proposed for rationalizing the amount of data that are collected for the calculation of cost. These include: limiting data collection to those services that healthcare providers deem are related to the illness or therapy under study, limiting data collection to medical services provided by the study site, and limiting the number of participants for whom economic data are collected. While all three have the strength of reducing resources required for data collection, most if not all have the cost of potentially undermining what can be learned from the economic evaluation.

Limit data collection to disease-related services

In cases where we do not expect to affect all medical service use, we may be concerned: (1) about the cost of collecting "unrelated" service use and (2) that the variability introduced by collecting "unrelated" service use may make it more difficult to detect differences that are expected from "related" service use. One proposed solution to this problem is to limit data collection to disease-related services.

Limiting data collection via this method may succeed in reducing both the cost of the trial and the variability that is introduced into our estimate of cost by collecting all services of certain types, independent of the reason for these services. However, there are several potential problems with this strategy. For example, there is little if any evidence about the accuracy, reliability, or validity of judgments about what is and what is not disease related. Much of medical practice is multifactorial. In a trial for heart failure, a participant with more severe heart failure may be hospitalized for a comorbid condition, whereas a

participant with the same comorbidity, but with milder heart failure may not be hospitalized. Answering counterfactuals such as "would the participant have been hospitalized had his or her heart failure been other than it is?" is the role of randomized trials, and it is unclear that provider judgment is any better at answering questions about the reasons for treatment than they are at answering questions about overall treatment efficacy.

In addition, for participants with complicated conditions, procedures ordered for the diagnosis and treatment of one condition may be complements or substitutes for procedures ordered for other conditions. Thus, even if providers are certain that they are ordering tests or procedures for the illness or therapy under study, it is possible that the same tests or procedures would have been ordered later for different reasons.

While these two limitations argue against collecting only that medical service use which is considered to be disease related, they do not argue against identifying those services that are or are not judged to be disease related. One can then perform a secondary analysis that evaluates cost differences that are stratified by these judgments.

Limit the delivery settings in which medical service use is collected

It is usually less costly to collect information about use of medical services when study participants see study personnel in study centers than it is when they see nonstudy providers in centers that are unaffiliated with the trial. Given the relatively higher cost of collecting data in the latter settings compared with the former, some may propose to eliminate collection of data from nonstudy providers and sites. However, ignoring the cost of nonstudy center care may lead to the conclusion that one treatment group has a higher cost than another, when in truth cost among the groups is similar, but incurred in different settings. Thus, if large amounts of care are provided in nonstudy settings, or if there are reasons to believe that utilization in these settings will differ between treatment groups, we should consider extending data collection efforts to these settings.

Presuming that the study is designed to collect data on service use from providers and in centers that are unaffiliated with the trial, a strategy should be in place to identify when study participants receive these services. This information may be easy to collect if administrative data such as insurance records are available for the study population or if they receive the preponderance of care from study investigators. In many cases, however, the data will have to come from—sometimes fallible—reporting by study participants or study participants' proxies.

A large literature has found the reliability of respondent recall to be mixed [37, 38]. It has been noted that accuracy declines the longer the recall interval

and is greater for some types of services (e.g., hospitalizations) than it is for others (e.g., outpatient services and medication use). For example, one study found that 90% of patients reported that they had been hospitalized when they were interviewed approximately 5 months after their discharge; another found that 50% reported they had been hospitalized 10–11 months after their admission [39]. A third study reported that patients reported 91% of ambulatory physician visits when they were interviewed 2 weeks after the visit, but reported only 30% when asked to recall visits during the previous year [40]. Accuracy also declines as the perceived level of importance of the service use decreases, the social acceptability of the disease decreases, and the mental impairment of the respondent increases.

If the data come from study participants or proxies, a decision should be made about how and how often study participants will be contacted. Methods include face-to-face interviews, telephone interviews [41, 42], Internet surveys [43], mail surveys [44, 45] and mobile app surveys. The more frequently participants are interviewed, the less likely the data will be affected by recall bias; however, the cost of data collection is higher and the burden on study participants and proxies is greater. The less frequently they are interviewed, the cost of the data is lower and the burden is lessened. The data, however, are more likely to be affected by recall bias. Clarke et al. have proposed formal methods for addressing this trade-off [46], while Seidl et al. try to avoid the problem by sampling service use data [47].

Even if study participants remember that they have seen a physician or have gone to a hospital, they often will not know what services they received. Options for identifying these services include contacting the care providers directly or assuming a standard set of services, such as the marginal cost for a usual visit to a family physician. Standard sets of services might be assigned differentially to account for differences by indication, for example, diagnosis-related group (DRG); differences in duration, for example, brief versus intermediate office visits; or the presence of an intensity marker such as a major diagnostic study.

When collecting data from study participants, another issue is whether to simply rely on their memory or whether to have each study participant use a memory aid, such as a diary [48]. Van den Brink and colleagues have reported similar results from both diaries and questionnaires [49]. Marques et al., on the other hand, have reported that use of diaries reduces missing data [50]. Diary forms can be actual pages in the case report form. More commonly, they are aids that are provided to the participant which are reviewed or abstracted by study personnel so the information can be coded on the case report form. If diaries are used, the investigators should decide how often to contact each study participant with a reminder to use the diary.

Lastly, who will collect these data: the research staff at the study site or a contract research organization? Contract research organizations can often provide interviewers and computerized interview scripts with built-in skip logic and online data entry. On the other hand, the study-site research staff sometimes are wary of having outside organizations contacting study participants, particularly if the participants have a stigmatizing disease. However, these same staff may be pleased to avoid having to interview participants about medical service use on a regular basis.

Limit participants from whom economic data are collected

A third method of reducing data collection cost is to limit the number of participants from whom these data are collected. If, however, there is the same or less power for the economic endpoints than there is for the clinical endpoints, such a strategy will limit the confidence we will have in the study's conclusions.

If there is more power for the economic than the clinical endpoints, we might choose to limit the number of participants for whom economic data are collected. In this case, the data should come from a random sample of all study participants. Nonrandom sampling, on the other hand, is problematic. For example, in an effort to speed completion of the trial, investigators might enroll participants who agree to participate in the trial but are unwilling to cooperate with the economic assessment. Such an approach can be problematic because self-selection by study participants places the validity of the economic study in jeopardy. The economic assessment could end up comparing an estimate of effects from the entire study population with an estimate of cost derived from a relatively small and nonrandom subset of the population.

Another approach to reducing data collection burden is to limit the economic assessment to the study centers that elect to participate. Unfortunately, this approach can also lead to difficulties. In one study that used voluntary data collection [51], the average length of stay (LOS) among study participants at centers that provided supplementary data was not the same as the average stay from nonresponding centers. This difference made the data analysis substantially more complex in that LOS in the intensive care unit could not be analyzed directly, but instead had to be imputed based on LOS in the hospital and other explanatory factors.

We have focused principally on the use and cost of medical services. Other types of costs that sometimes are documented within economic evaluations include direct nonmedical costs such as transportation, time costs associated with either illness or treatment, and intangible costs. The types of costs that should be included in an analysis depend on what is affected by illness and its

treatment and what is of interest to decision-makers. For example, the National Institute for Health and Care Excellence (UK) and the Australian Pharmaceutical Benefits Scheme have indicated they have little interest in time costs.

Work loss, a time cost, is incurred both when workers miss time from work (absenteeism) and when they are at work but their illness makes them less productive than they would have been had they not been ill (presenteeism) [52]. A large number of published instruments have been developed to assess absenteeism and presenteeism, including the Work Productivity and Activity Impairment Questionnaire (WPAI) [53], the Health and Work Performance Questionnaire (HPQ) [54], and the Work Limitations Questionnaire (WLQ) [55]. All three instruments have been used in the evaluation of therapies for a broad range of illnesses, but the WPAI has been used and assessed most extensively [56, 57]. A number of authors have reviewed these and other instruments [56–58]. One finding is that choice of instruments matters because they can differ widely in their estimates [59]. Finally, much of the validation of these instruments has occurred in settings such as call rooms and among production line workers. It is less clear that they perform well in the valuation of work that requires creative thinking or team work or that need not be performed at a desk in an office [60, 61].

In summary, the general strategy we recommend is to identify types of medical services that will be documented within the case report form (e.g., hospitalization) and then to collect all of these services independently of their cause. As in the gold standard evaluation, this strategy allows randomization to determine those services that are and are not related to the study therapy. Unless we have an a priori hypothesis that differences will be found among low-cost medical services, we recommend that attention be given to high-cost services that are likely to make up a large portion of total cost. For example, for heart failure, we might focus on hospitalization—for severe heart failure we might also focus on outpatient visit cost; for hospitalized infections, we might focus on intensive, intermediate, and routine care unit LOS and major procedures; and for asthma, we might focus on hospitalizations, emergency department visits, and co-medications.

During preplanning, we can develop methods for reducing data collection cost, without undermining the validity of the results, by identifying those services that are likely to make up a large portion of the difference in cost between the treatment groups. If we determine that the therapy is likely to affect the number of hospitalizations, collect information that will provide a reliable estimate of the cost of these hospitalizations. If it is likely to affect days in the hospital and location in the hospital, collect this information.

If the therapy is principally likely to affect outpatient care, collect measures of outpatient care, and the like.

2.3.3 At what level should medical service use be aggregated?

A strategy commonly adopted for calculating cost in trials is to count medical service use that occurs during the trial and to multiply this count by a set of price weight estimates for these services. The level at which medical services are aggregated—for example, for inpatient care, should we count hospitalizations? Days in the hospital stratified by location in the hospital? Should we in addition count individual services provided in the hospital?—and the resulting price weight estimates required for the calculation of cost, depend on a number of factors, including whether we expect the intervention to affect the number of hospitalizations that occur, the LOS of a hospitalization when it occurs, or the intensity of medical services utilized during the stay. In making decisions about the level of aggregation at which the data will be collected, the investigator should consider the likely difference more or less aggregated information will have on the study result as well as the cost of collecting more or less aggregated data.

At the most aggregate level, outpatient care can be recorded as the number of visits. Alternatively, diagnostic tests, procedures, and treatments can be recorded as well. For hospital care, the types of services that are counted—and the resulting price weight estimates required for the calculation of cost—often depend on the setting in which the therapies under investigation are expected to be used. For therapies used predominantly in hospital settings, a common approach is to sum the individual costs of a hospital stay, such as those associated with days in the hospital, stratified by intensity of care, laboratory evaluations, procedures, and medications [51, 62–64].

For therapies used predominantly in outpatient settings, it is more common to collect information about hospital diagnoses and LOS. These hospitalizations can be valued by use of aggregate measures of hospital cost, such as diagnosis-related group (DRG) payments [64–69] or an estimate of the cost per day multiplied by the number of days in the hospital [70–72]. When this latter strategy is adopted, different studies may use cost estimates with varying levels of specificity. For example, one of the least specific approaches would be to use a single cost estimate from a single center to value all hospitalizations at all centers, whereas one of the most specific would be to use diagnosis-specific price weight estimates from each center that participated in the study. Most studies adopt a strategy that falls somewhere between these bounds.

2.3.4 Which price weight estimates should be used for the study?

Sources of price weights differ by country. For example, in the US, there are a number of publicly available sources, including hospital charges adjusted using cost-to-charge ratios, data from internal hospital costing systems, DRG payments for hospitalizations [73], resource-based relative value units for physician services [74, 75], and fee schedules for laboratory tests, medications, and durable medical equipment.

Readily available data in Europe and in some other parts of the world may include, depending on the country: fee schedules, data from DRG studies, and data from hospital costing systems. In some cases, data have been obtained from a limited number of centers which have cost accounting systems or from administrative databases that have been developed in some countries [76].

It may also be possible to develop trial-specific costing exercises, for example, analysis of accounting data [77, 78], time and motion studies [79], or cost allocation projections for expensive experimental capital equipment [80].

Detailed discussion related to the development of price weights for the evaluation is provided in Chapter 3.

2.3.5 How naturalistic should the study design be?

Given that the primary purpose of cost-effectiveness analysis is to inform real-world decision-makers about how to respond to real-world healthcare needs, the more naturalistic the trial, in terms of participants, analysis based on the intention to treat, and limitation of loss to follow-up, the more likely the data developed within the trial will speak directly to the decision question. Often, however, trials adopt less naturalistic study designs or do not actively seek to limit loss to follow-up.

Sample inclusion criteria

First, many phase III efficacy trials and some effectiveness trials employ stringent inclusion and exclusion criteria, and may employ cream skimming—the selection of a subset of the healthiest eligible patients—or reverse cream skimming—the selection of a subset of the sickest eligible patients—in the construction of their study sample. The resulting sample may not represent the more heterogeneous population found in general practice that decision-makers consider when making resource allocation decisions. The efficacy, effectiveness, and efficiency of therapies may be different in the more homogeneous set of participants included in the trial as compared to patients in the general population, and—the arguments of decision analysts to the contrary—it is not clear that valid and reliable data are available to enable translation from what was

observed in the homogenous sample to what would have been observed in the heterogeneous one.

Intention-to-treat analysis

Second, given that in real-world settings, economic questions relate to treatment decisions, for example, whether to prescribe a therapy, not whether the patient received the therapy prescribed nor whether, once they started the prescribed therapy, they were switched to other therapies, cost and benefit associated with these later decisions should be attributed to the initial treatment decision. Thus, trial-based cost-effectiveness analyses should adopt an intention-to-treat design.

Loss to follow-up

Third, we should design studies in such a way that they minimize the occurrence of missing and censored data. For example, study designs should include plans to aggressively pursue participants and data throughout the trial. One long-term study of treatment for bipolar disorder was designed from the outset to respond to missed interviews by: (1) intensive outreach to reschedule the assessment followed by (2) telephone assessment followed by (3) interview of a proxy who had been identified and consented at the time of randomization [81].

Investigators should also ensure that follow-up continues until the end of the study period. As Polsky et al. [82] have argued, data collection should not be discontinued simply because a participant reaches a clinical or treatment stage such as failure to respond (as, for example, happens in some antibiotic, cancer chemotherapy, and psychiatric drug trials). This last recommendation may conflict with some commonly used efficacy designs that are event driven and end follow-up of individual participants when they reach such a stage. However, given that failure often is associated with a change in the pattern of cost, for example, due to the initiation of alternative therapies and the potential elongation of the duration of the episode of care or due to the discontinuation of all treatments, discontinuation of these participants from the economic study is likely to bias the results of an economic evaluation that is conducted as part of the trial.

In general, bias related to follow-up is least likely to occur if all study participants are followed for a fixed time period, for example, all study participants are followed for 1 year after randomization or until death, whichever is sooner. Another design which is unbiased—but, as discussed in Chapter 6, requires the use of analytic methods for addressing censored data—is to end follow-up on a given calendar date, for example, when the 500th death is observed in the trial. This design, which has been used in a number of long-term

cardiovascular trials [83, 84], gives rise to what is called "administrative censoring" and the resulting data are usually classified as being "missing completely at random" [85].

Protocol-induced cost and effect

Finally, clinical trial protocols often attempt to standardize the care of participants in trials, in which case, the care delivered may differ from usual care. Trial protocols may require a substantial number of investigations and diagnostic tests that would not be performed under normal clinical practice. They may also prescribe aggressive documentation and treatment of the outcomes observed in trials.

A common reaction is to exclude the cost of protocol-induced services from the economic evaluation, in part because it is argued that they would not have been provided in usual practice. However, inclusion or exclusion of services should not be based on their likelihood of use in usual practice. It should instead be based on whether the services could affect outcomes—and thus cost and effect—observed in the trial. "Cadillac" care may be associated with "Cadillac" outcomes, and excluding one but not the other is likely to bias the results of the trial.

A service such as collection of genetic samples that will not be analyzed until trial follow-up is completed can easily be excluded from the analysis. Exclusion of services whose results are provided directly to treating clinicians, on the other hand, is more problematic. If these services are thought to have little or no impact on outcome, they will usually represent the simple addition of a constant to our cost estimates. Thus, while their inclusion or exclusion affects the arithmetic mean cost for each treatment group, it will have little effect on the difference between the costs of therapies.

If, instead, the services are thought to have an impact, and if in usual practice they would have been performed in some participants, but not in others, the fact that in the trial they are mandated for all participants will tend to bias the cost analysis to the null. In this case, we can no longer observe differences in these services that might have existed between the groups. Unlike the first example, inclusion or exclusion of these costs makes little difference not because they represent a constant, but because neither option can correct the bias that may have been induced by the protocol.

Protocol-induced testing can have a number of other pernicious effects. It can lead to the avoidance of outcomes that would have occurred in usual practice but in the trial are avoided by actions in response to the testing. In addition, it may lead to the trial's detection and treatment of outcomes earlier in the disease process than would have been the case in usual practice. It may also result in the

trial's detection and treatment of outcomes that would never have been detected and treated in usual care [86]. In these cases, it may be difficult to determine how many outcomes would have been detected in usual practice, what the cost of care would have been, and whether the therapy tended to avoid the average outcome, those outcomes that in usual practice would never have been detected, or those outcomes that would have been most severe and caused the greatest harm in terms of cost and effect.

Protocol-mandated treatment may also cloud what we can say about the cost and effect of therapy, because its impact depends on whether the mandated treatments are cost-effective or not. If protocol-mandated treatment is itself cost-effective, its inclusion in the protocol will benefit the therapy that is associated with the larger number of treated outcomes. That is because this therapy's relative cost and effect otherwise would have been less favorable than they are given the protocol-mandated cost-effective services. If, on the other hand, the mandated treatment is not cost-effective, its inclusion in the protocol will disadvantage the therapy that is associated with the larger number of treated outcomes because its relative cost and effect otherwise would have been more favorable than they are given the protocol-mandated cost-ineffective services.

2.3.6 What should we do if the full benefit and cost will not be observed during the trial?

Most trials that evaluate therapies for chronic conditions end study follow-up before the study medication would be discontinued in usual practice, the latter of which often occurs when the patient dies. In cases where the trial ends follow-up before a therapy would typically have been discontinued, it is good practice to evaluate the cost and effect that were observed during the trial. In such a "within-trial" evaluation, we should maintain the same time horizons for cost and effect that were observed in the trial. For example, if follow-up for the trial was for 1 year, then cost and effect should be measured for 1 year.

If, however, long-term use yields outcomes that generally cannot be observed during the shorter time frames employed in trials, or if the cost-effectiveness ratio is heterogeneous with time of follow-up, making therapeutic decisions based solely on results observed within short-term trials may be inappropriate. To address these limitations, a number of investigators have developed decision analytic models that use data from the trial and, in some cases, data from clinical registries and other sources, to attempt to address this issue [64, 69]. (Note: as with the "within-trial" evaluation, we should maintain the same time horizons for the projections of cost and effect.) At a minimum, for trials that end follow-up before therapy would have ended, we need to make plans for credibly assessing longer-term cost and effect.

Table 2.1 Cost-effectiveness evaluated by years of follow-up within the trial and by projected years[a]

Years of follow-up	Point estimate	95% CI
Within the trial		
1	Dominated	168,884 to Dominated
2	282,857	45,577 to Dominated
3	73,529	Dominates to Dominated
4	12,074	Dominates to Dominated
5	15,258	Dominates to 122,772
Longer-term projection		
10	12,246	Dominates to 42,263
15	8578	Dominates to 26,721
20	7320	681 to 21,841

[a] Authors' unpublished data.

One potential advantage of developing decision analytic models directly from trials with multiple years of follow-up is that we can evaluate the trajectory of the cost-effectiveness ratio over time. For example, in a 5-year trial, we can evaluate cost-effectiveness after each year of observation and then make projections for the first 5 years after the trial, the first 10, etc. Such a projection is provided in Table 2.1.

Table 2.1 indicates that after 1 year, the therapy had not demonstrated value for the cost. As the period of follow-up within the trial increased, the evidence of value grew, although even after 5 years we could not rule out an upper limit for the confidence interval of the cost-effectiveness ratio of 123,000. Projection of results provides two important findings. First, at 5 years we have a reasonable approximation of the point estimate for the ratio that we project would be observed with longer follow-up. In other words, the ratio at 5 years is relatively homogeneous with projected results for longer periods of follow-up. Thus, we do not need to make a 40-year projection before we understand the likely point estimate for this therapy. Second, even after a relatively short projection, our results suggest that the confidence interval will narrow to a range where we can be confident of the value for the cost. Given the "model" uncertainty that surrounds 30- and 40-year projections, observing good value after only a short period of projection allows more confidence in our result.

Another advantage of developing a decision analytic model from the data from a trial is that the trial enables a limited assessment of the internal validity of the model. In other words, we are able to determine if the model can replicate

the cost-effectiveness trajectory observed in the trial. Of course good internal validity data says nothing about the validity of the projection.

When is it inappropriate to perform an economic evaluation within a trial?

The presence or absence of an economic advantage should not be a deciding factor on whether to conduct an economic evaluation within a clinical trial, because both results provide useful information. Rather, as indicated in the section on naturalism of the trial, the most common reasons why it would be inappropriate to perform such an evaluation are when the trial design is such that no unbiased evidence about economic value will be observable (even non-significantly) during the period of observation. For example, if we have mandated so much care that we overwhelm all potential differences in cost, or if we have differentially followed participants based on outcome, it is unlikely that an economic assessment will be informative.

Practically, investigators may withhold performance of an economic evaluation if they believe that it will not affect the decision to use the therapy. For example, a therapy may be so effective that people will not worry about its cost, for instance, therapy that adds a year of life expectancy for a disease that is highly fatal and was previously untreatable. Similarly, in some healthcare systems, therapies may be so novel that clinicians will use them even if they are not cost-effective.

What contributes to a successful economic evaluation within a trial?

Successful economic assessments conducted as part of clinical trials require a commitment to carry out such assessments. This commitment is characterized by early planning of the economic component of the trial, alerting the clinical investigators at the outset that the economic data will be collected along with the clinical data, and expecting all participants in the study to contribute both clinical and economic data to the study. In contrast, less successful economic assessments tend to have the clinical study designed independently and in advance of the economic study, the clinical investigators are recruited before the economic study is in place, and the economic study is introduced just as the trial initiates enrollment.

Trials least successful at recruiting participants are often ones in which the implicit message from the organizers is that the investigators are not obligated to enroll participants into the economic assessment. Trials should be designed so that enrolled participants are required to contribute both clinical and economic data to the study. One way to do so is to integrate economic data collection into the trial's case report forms, rather than segregating them so that economic data collection appears like a separate activity. Some may be concerned that such a

requirement raises ethical issues, but trial exclusion criteria routinely eliminate patients who are unlikely to cooperate with data collection. We can minimize the chance that a patient believes she can participate in the trial and contribute clinical data but not economic data if we use a single informed consent agreement for both the clinical and economic information. If we seek to obtain billing data from providers who do not participate in the study, we should also ask study participants to sign forms that release their billing information.

Successful economic assessments as part of clinical trials also require cooperation and coordination among everyone involved in the trial, including organizers, clinical researchers, and study participants. Disinterest or dissatisfaction among any of these groups can lead to collection of flawed data and can undermine the study. They also require a willingness to take the risk of determining that the therapy may be clinically effective yet not cost-effective.

2.4 **Discussion**

Many opportunities exist for the incorporation of economic assessments into the randomized trials used to assess medical therapies. When these assessments are conducted during the drug development process, they provide data about a drug's value early in its product life, and these data can be used by policymakers, drug manufacturers, healthcare providers, and patients when the therapy is first introduced in the market. Better data about a therapy's economic effect potentially can be collected once it is used widely, but for making an early decision, data from phase III studies (or their equivalent in nondrug trials) are the best that are available.

In this chapter, we have presented a broad overview of some of the current practices and methodological issues that are part of an economic assessment. However, a wide diversity exists in the design and implementation of economic assessments. Some who organize economic assessments limit participation to study participants for whom large amounts of computerized billing data are available. Others focus on collecting hospitalization data in the form of electronic records from participating centers, thereby avoiding the need to deal with paper records. Another group of organizers prefers to collect the economic data from case report forms. This last option may be particularly useful for multinational trials, where the availability of electronic data may differ from country to country.

Economic analyses conducted within randomized trials share many design issues with traditional clinical trials, but they have unique issues as well. Means for addressing the latter set of issues continue to undergo rapid development, and the methods for doing these studies continue to evolve. We address a number of the latter issues in succeeding chapters.

References

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

2 **Garber AM.** Advances in cost-effectiveness of health interventions. In *Handbook of Health Economics Volume 1A*, edited by AJ Culyer and JP Newhouse. Amsterdam: North-Holland, 2000, 181–221.

3 **Eisenberg JM.** Clinical economics. A guide to the economic analysis of clinical practice. *JAMA*. 1989; **262**: 2879–86.

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

5 **Gold MR, Siegel JE, Russell LB, Weinstein MC (eds).** *Cost-Effectiveness in Health and Medicine*. Oxford: Oxford University Press, 1996.

6 **Hornbrook MC.** Definition and measurement of episodes of care in clinical and economic studies. In *Conference Proceedings: Cost Analysis Methodology for Clinical Practice Guidelines*, edited by MC Grady and KA Weis. DHHS Pub. No. (PHS) 95-001. Rockville, MD: Agency for Health Care Policy and Research, 1995, 15–40.

7 **Keeler EB.** Using episodes of care in effectiveness research. In *Medical Effectiveness Research Data Methods*, edited by ML Grady and HA Schwartz. DHHS Pub. No (PHS) 92–0056. Rockville, MD: Agency for Health Care Policy and Research, 1992, 172–6.

8 **Steinwachs DM.** Episode of care framework: utility for medical effectiveness research. In *Medical Effectiveness Research Data Methods*, edited by ML Grady and HA Schwartz. DHHS Pub. No (PHS) 92–0056. Rockville, MD: Agency for Health Care Policy and Research, 1992, 161–71.

9 **Hornbrook MC, Hurtado AV, Johnson RE.** Health care episodes: definition, measurement, and use. *Med Care Rev*. 1985; **42**: 163–218.

10 **Brainsky A, Glick H, Lydick E, Epstein R, Fox KM, Hawkes W, Kashner TM, Zimmerman ST, Magaziner J.** The economic cost of hip fractures in community-dwelling older adults: a prospective study. *JAGS*. 1997; **45**: 281–7.

11 **Mehta SS, Suzuki S, Glick HA, Schulman KA.** Determining an episode of care using claims data. Diabetic foot ulcer. *Diabetes Care*. 1999; **22**: 1110–15.

12 **Schulman KA, Yabroff KR, Kong J, Gold KF, Rubenstein LE, Epstein AJ, Glick H.** A claims data approach to defining an episode of care. *Health Serv Res*. 1999; **34**: 603–21.

13 **Glick HA, Polsky D, Schulman K.** Trial-based economic evaluations: an overview of design and analysis. In *Economic Evaluation in Health Care. Merging Theory and Practice*, edited by M Drummond and A McGuire. Oxford: Oxford University Press, 2001, 113–40.

14 **Glick H, Kinosian B, Schulman K.** Decision analytic modelling: some uses in the evaluation of new pharmaceuticals. *Drug Inform J*. 1994; **28**: 691–707.

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

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

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

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

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

20 **Hollingworth W, McKell-Rddwood D, Hampson L, Metcalfe C.** Cost-utility analysis conducted alongside randomized controlled trials: are economic end points considered in sample size calculations and does it matter. *Clin Trials*. 2013; **10**: 43–53.

21 **Glick HA.** Sample size and power for cost-effectiveness analysis (part 1). *Pharmacoeconomics*. 2011; **29**: 189–98.

22 **Koerkamp BG, Spronk S, Stijnen T, Hunink MGM.** Value of information analyses of economic randomized controlled trials: the treatment of intermittent claudication. *Value Health*. 2010; **13**: 242–50.

23 **Brown M, Glick HA, Harrell F, Herndon J, McCabe M, Moinpour C, Schulman KA, Smith T, Weeks J, Seils DM.** Integrating economic analysis into cancer clinical trials: the National Cancer Institute-American Society of Clinical Oncology economics workbook. *J Natl Cancer Inst Monogr*. 1988; **24**: 1–28.

24 **Towse A.** What is NICE's threshold? An external view. In *Cost Effectiveness Thresholds: Economic and Ethical Issues*, edited by N Devlin, A Towse, London: Kings Fund/Office for Health Economics, 2002, 25–30.

25 **Devlin N, Parkin D.** Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Econ*. 2004; **13**: 437–52.

26 **George B, Harris A, Mitchell A.** Cost-effectiveness analysis and the consistency of decision making. Evidence from pharmaceutical reimbursement in Australia (1991 to 1996). *Pharmacoeconomics*. 2001; **19**: 1103–9.

27 **Hirth RA, Chernew ME, Miller E, Fendrick M, Weissert WG.** Willingness to pay for a quality-adjusted life year: in search of a standard. *Med Decis Making*. 2000; **20**: 332–42.

28 **Towse A.** Should NICE's threshold range for cost per QALY be raised? *BMJ*. 2009; **338**: b181.

29 *The Patient Protection and Affordable Care Act*. PL 111-148. 23 March 2010.

30 **Gupta AK.** Pharmacoeconomic analysis of ciclopox nail lacquer solution 8% and the new oral antifungal agents used to treat dermatophyte toe onychomycosis in the United States. *J Am Acad Dermatol*. 2000; **43**: S81–S95.

31 **Chisholm D, Knapp KMJ, Knudsen HC, Amaddeo F, Gaite L, van Wijngaarden B, Epsilon Study Group.** Client Socio-Demographic and Service Receipt Inventory-European version. *Br J Psychiatry Suppl*. 2000; **39**: S28–S33.

32 **Wimo A, Norberg G.** Validity and reliability of assessments of time. Comparisons of direct observations and estimates of time by the use of the resource utilization in dementia (RUD)-instrument. *Arch Gerontol Geriatr*. 2007; **44**: 71–81.

33 **Sano M, Zhu CW, Whitehouse PJ, Edland S, Jin S, Ernstrom K, Thomas RG, Thal LJ, Ferris S.** ADCS Prevention Instrument Project: pharmacoeconomics: assessing health-related resource use among healthy elderly. *Alzheimer Dis Assoc Disord*. 2006; **20**: S191–S202.

34 **Chernyak N, Ernsting C, Icks A.** Pre-test of questions on health-related resource use and expenditure, using behaviour coding and cognitive interviewing techniques. *BMC Health Serv Res*. 2012; **12**: 303.

35 **Ridyard CH, Hughes DA.** Development of a database of instruments for resource-use measurement: purpose, feasibility, and design. *Value Health*. 2012; **15**: 650–5.

36 **Ridyard CH, Hughes DA.** Methods for the collection of resource use data within clinical trials: a systematic review of studies funded by the UK Health Technology Assessment Program. *Value Health*. 2010; **13**: 867–72.

37 **Bhandari A, Wagner T.** Self-reported utilization of health care services: improving measurement and accuracy. *Med Care Res Rev*. 2006; **63**: 217–35.

38 **Evans C, Crawford B.** Patient self-reports in pharmacoeconomic studies. *Pharmacoeconomics*. 1999; **15**: 241–56.

39 **Cannell CF.** *A Summary of Research Studies of Interviewing Methodology, 1959–1970*. Vital and Health Statistics: Series 2, Data Evaluation and Methods Research; Number 69, DHEW Publication; Number (HRA) 77–1343. Washington, DC: US Government Printing Office, 1977.

40 **Roberts RO, Bergstralh EJ, Schmidt L, Jacobsen SJ.** Comparison of self-reported and medical record health care utilization measures. *J Clin Epidemiol*. 1996; **49**: 989–95.

41 **Unger WJ, Coyte PC.** Health services utilization reporting in respiratory patients. *J Clin Epidemiol*. 1998; **51**: 1335–42.

42 **Brown JB, Adams ME.** Patients as reliable reporters of medical care process. *Med Care*. 1992; **30**: 400–13.

43 **Palmer L, Johnston SS, Rousculp MD, Chu BC, Nichol KL, Mahadevia PJ.** Agreement between internet-based self- and proxy-reported health care resource utilization and administrative health care claims. *Value Health*. 2012; **15**: 458–65.

44 **Patel A, Moran P, Leese M, Mann A, Knapp M.** A comparison of two methods of collecting economic data in primary care. *Fam Pract*. 2005; **22**: 323–7.

45 **Sahqvist S, Song Y, Bull F, Adams E, Preston J, Ogilvie D.** Effect of questionnaire length, personalization and reminder type on response rate to a complex postal survey: randomized controlled trial. *BMC Med Res Methodol*. 2011; **11**: 62.

46 **Clarke PM, Fiebig DG, Gerdtham UG.** Optimal recall length in survey design. *J Health Econ*. 2008; **27**: 1275–84.

47 **Seidl H, Meisinger C, Wende R.** Empirical analysis shows reduced cost data collection may be an efficient method in economic clinical trials. *BMC Health Serv Res*. 2012; **12**: 318.

48 **Goossens ME, Rutten-van Molken MP, Vlaeyen JW, van der Linden SM.** The cost diary: a method to measure direct and indirect costs in cost-effectiveness research. *J Clin Epidemiol*. 2000; **53**: 688–95.

49 **Van den Brink M, van den Hout WB, Stiggelbout AM, Putter H, van de Velde CJH, Lievit J.** Self-reports of health-care utilization: diary or questionnaire. *Int J Technol Assess Health Care*. 2005; **21**: 298–304.

50 **Marques E, Johnson EC, Gooberman-Hill R, Blom AW, Noble S.** Using resource use logs to reduce the amount of missing data in economic evaluations alongside trials. *Value Health*. 2013; **16**: 195–201.

51 **Glick H, Willke R, Polsky D, Llana T, Alves WM, Kassell N, Schulman K.** Economic analysis of tirilazad mesylate for aneurysmal subarachnoid hemorrhage: economic evaluation of a Phase III clinical trial in Europe and Australia. *Int J Tech Assess Health Care*. 1998; **14**: 145–60.

52 **Burton WN, Conti DJ, Chen CY, Schultz AB, Edington DW.** The role of health risk factors and disease on worker productivity. *J Occup Environ Med*. 1999; **41**: 863–77.

53 **Reilly MC, Zbrozek AS, Duke EM.** The validity and reliability of a Work Productivity and Activity Impairment instrument. *Pharmacoeconomics*. 1993; **4**: 355–65.

54 **Kessler RC, Barber C, Beck A, Berglund P, Cleary PD, McKenas D, Pronk N, Simon G, Strang P, Ustun TB, Wang P.** The World Health Organization Heath and Work Performance Questionnaire (HPQ). *J Occup Environ Med*. 2003; **45**: 156–74.

55 **Lerner D, Amick BC, Rogers WH, Malspeis S, Bungay K, Cynn D.** The Work Limitations Questionnaire. *Med Care*. 2001; **39**: 72–85.

56 **Prasad M, Wahlqvist P, Shikiar R, Shih YCT.** A review of self-reported instruments measuring health-related work productivity. *Pharmacoeconomics*. 2004; **22**: 225–44.

57 **Mattke S, Balakrishnan A, Bergamo G, Newberry SJ.** A review of methods to measure health-related productivity loss. *Am J Manag Care*. 2007; **13**: 211–17.

58 **Beaton DE, Tang K, Gignac MAM, Lacaille D, Badley EM, Anis AH, Bombardier C.** Reliability, validity, and responsiveness of five at-work productivity measures in patients with rheumatoid arthritis and osteoarthritis. *Arthritis Care Res*. 2010; **62**: 28–37.

59 **Zhang W, Gignac WAM, Beaton D, Tang K, Anis AH.** Productivity loss due to presenteeism among patients with arthritis: estimates for 4 instruments. *J Rheumatol*. 2010; **37**: 1805–14.

60 **Nicholson S, Pauly MV, Sharda C, Szreck H, Berger ML.** Measuring the effects of work loss on productivity with team production. *Health Econ*. 2006; **15**: 111–23.

61 **Pauly MV, Nicholoson S, Polsky D, Berger ML, Sharda C.** Valuing reductions in on-the-job illness: 'presenteeism' from managerial and economic perspectives. *Health Econ*. 2008; **17**: 469–85.

62 **Legendre C, Norman D, Keating M, Maclaine G, Grant D.** Valaciclovir prophylaxis of cytomegalovirus infection and disease in renal transplantation: an economic evaluation. *Transplantation*. 2000; **70**: 1463–8.

63 **Schulman KA, Dorsainvil D, Yabroff KR, DiCesare J, Felser J, Eisenberg JM, Glick HA.** Prospective economic evaluation accompanying a trial of GM-CSF/IL-3 in patients undergoing autologous bone-marrow transplantation for Hodgkin's and non-Hodgkin's lymphoma. *Bone Marrow Transplant*. 1998; **21**: 607–14.

64 **Mark DB, Hlatky MA, Califf RM, Naylor CD, Lee KL, Armstrong PW, Barbash G, White H, Simoons ML, Nelson CL.** Cost effectiveness of thrombolytic therapy with tissue plasminogen activator as compared with streptokinase for acute myocardial infarction. *N Engl J Med*. 1995; **332**: 1418–24.

65 **Mauskopf J, Cates S, Griffin A, Neighbors D, Lamb S, Rutherford C.** Cost effectiveness of zanamivir for the treatment of influenza in a high risk population in Australia. *Pharmacoeconomics*. 2000; **17**: 611–20.

66 **Markowitz M, Mauskopf J, Halpern M.** Cost-effectiveness model of adjunctive lamotrigine for the treatment of epilepsy. *Neurology*. 1998; **51**: 1026–33.

67 **Riviere M, Wang S, Leclerc C, Fitzsimon C, Tretiak R.** Cost-effectiveness of simvastatin in the secondary prevention of coronary artery disease in Canada. *CMAJ*. 1997; **156**: 991–7.

68 **Johannesson M, Jonsson B, Kjekshus J, Olsson A, Pederson T, Wede H.** Cost effectiveness of simvastatin treatment to lower cholesterol levels in patients with coronary heart disease. Scandinavian Simvastatin Study Group. *N Engl J Med*. 1997; **336**: 332–6.

69 **Glick H, Cook J, Kinosian B, Pitt B, Bourassa MG, Pouleur H, Gerth W.** Costs and effects of enalapril therapy in patients with symptomatic heart failure: an economic analysis of the Studies of Left Ventricular Dysfunction (SOLVD) treatment trial. *J Cardiac Fail*. 1995; **1**: 371–80.

70 **Mauskopf J, Richter A, Annemans L, Maclaine G.** Cost-effectiveness model of cytomegalovirus management strategies in renal transplantation. Comparing valaciclovir prophylaxis with current practice. *Pharmacoeconomics.* 2000; **18**: 239–51.

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

72 **Glick HA, Orzol SM, Tooley JF, Remme WJ, Sasayama S, Pitt B.** Economic evaluation of the Randomized Aldactone Evaluation Study (RALES): treatment of patients with severe heart failure. *Cardiovasc Drugs Ther.* 2002; **16**: 53–9.

73 Department of Health and Human Services, Health Care Financing Administration. Medicare program: hospital inpatient prospective payment systems. *Federal Register*. 1991; **56**: 4821.

74 Department of Health and Human Services, Health Care Financing Administration. Medicare program: fee schedule for physicians' services: final rule. *Federal Register*. 1991; **56**: 59502–811.

75 **Hsiao WC, Braun P, Dunn DL, Becker ER, Yntema D, Verrilli DK, Stamenovic E, Chen SP.** An overview of the development and refinement of the resource-based relative value scale. *Med Care*. 1992; **30**(suppl): NS1–NS12.

76 **Schulman K, Burke J, Drummond M, Davies L, Carlsson P, Gruger J, Harris A, Lucioni C, Gisbert R, Llana T, Tom E, Bloom B, Willke R, Glick H.** Resource costing for multinational neurologic clinical trials: methods and results. *Health Econ*. 1998; **7**: 629–38.

77 **Bridges M, Jacobs P.** Obtaining estimates of marginal cost by DRG. *Healthc Financ Manage*. 1986; **10**: 40–6.

78 **Granneman TW, Brown RS, Pauly MV.** Estimating hospital costs. A multiple-output analysis. *J Health Econ*. 1986; **5**: 107–27.

79 **Finkler SA, Knickman JR, Hendrickson G, Lipkin M, Thompson WG.** A comparison of work-sampling and time-and-motion techniques for studies in health services research. *Health Serv Res*. 1993; **28**: 577–97.

80 **Finkler SA.** Cost finding for high-technology, high-cost services: current practice and a possible alternative. *Health Care Manage Rev*. 1980; **5**: 17–29.

81 **Bauer MS, Williford WO, Dawson EE, Akiskal HS, Altshuler L, Fye C, Gelenberg A, Glick H, Kinosian B, Sajatovic M.** Principles of effectiveness trials and their implementation in VA Cooperative Study #430: 'Reducing the efficacy-effectiveness gap in bipolar disorder'. *J Affect Disord*. 2001; **67**: 61–78.

82 **Polsky D, Doshi JA, Bauer MS, Glick HA.** Clinical trial-based cost-effectiveness analyses of antipsychotics. *Am J Psychiatry*. 2006; **163**: 2047–56.

83 The SOLVD Investigators. Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. *N Engl J Med*. 1991; **325**: 293–302.

84 Scandinavian Simvastatin Survival Study Group. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). *Lancet*. 1994; **344**: 1383–9.

85 **Rubin DB.** Inference and missing data. *Biometrika*. 1976; **63**: 581–92.

86 **Eisenberg JM, Glick H, Koffer H.** Pharmacoeconomics: economic evaluation of pharmaceuticals. In *Pharmacoepidemiology*, edited by B Strom. New York: Churchill Livingstone, 1989, 325–50.