Office of Health Economics Research

Health Economics: An Introduction to Economic Evaluation

Third Edition

Gisela Kobelt

Health Economics: An Introduction to Economic Evaluation

Third Edition

Gisela Kobelt

2013

Office of Health Economics
Research

Office of Health Economics
Southside, 7th Floor
105 Victoria Street
London SW1E 6QT
United Kingdom
www.ohe.org

Printed in the United Kingdom
ISBN 978-1-899040-44-5

This monograph has undergone a rigorous peer review by the independent OHE Editorial Board and other experts in the field. The views expressed in this publication are those of the author and do not necessarily represent those of OHE.

About the Office of Health Economics

Founded in 1962, the OHE's terms of reference are to:

- Commission and undertake research on the economics of health and health care
- Collect and analyse health and health care data for the UK and other countries
- Disseminate the results of this work and stimulate discussion of them and their policy implications.

The OHE's work is supported by research grants and consultancy revenues from a wide range of UK and international sources.

About the Author

Gisela Kobelt is Managing Director of European Health Economics France, a company specialising in economic evaluation of health interventions and training courses in health economics and economic evaluation.

Until the end of 1997, Gisela was Vice President, Corporate Health Economics, at the Pharmacia & Upjohn Corporate Management Center in Stockholm and London. Prior to her responsibilities with Pharmacia, she created and headed the health economics department at Sandoz in Basel. Since 2007, has been a visiting professor at the Department of Orthopedics/Rheumatology at the University of Lund, Sweden.

Gisela holds a master's degree from the University of Strasbourg, an MBA from the Institute for Management Development in Lausanne, and a PhD in health economics from the Karolinska Institute.

Author of over 90 scientific publications in the field of health economics, several book chapters and numerous reports, Gisela serves on several editorial boards.

Acknowledgements

I am grateful to Professor Bengt Jönsson of the Stockholm School of Economics and to the OHE reviewers for their advice and helpful comments on the manuscript.
— *Gisela Kobelt*

FOREWORD

The Office of Health Economics is pleased to offer the third edition of this valuable guide to health economics.

The second edition of this publication appeared in 2002. In the decade since, the use of economic evaluation of new medical technologies as a basis for decisions about access to and reimbursement of medicines and medical services has expanded to an increasing number of countries and types of technology. At the same time, the methods themselves have evolved in response to experience and to changes in the ability to capture and analyse data. This new edition reflects those changes.

This book presents a comprehensive overview of approaches to health economic evaluation, illustrated throughout with examples and with guidance about what methods are appropriate in which situations. Written in an accessible style, the book offers important background both for those who will undertake evaluations and those who will use them as the bases for decisions. The author, Gisela Kobelt, has extensive experience in economic evaluation, making her perspective particularly insightful.

Professor Adrian Towse, Director
Office of Health Economics

CONTENTS

ILLUSTRATIONS

STUDY EXAMPLES

Chapter 1

HEALTH ECONOMICS: GENERAL ISSUES

The Economics of Health and Health Care

Over the past decades, the ability to provide treatment for an increasingly wide range of diseases has increased exponentially with the introduction of new technologies. Demand for care also has increased, partly in response to this, but also for other reasons. The resulting rise in health care costs has put considerable strain on finite resources, a situation that has worsened in the face of the current global economic slowdown.

Economic issues in health care are now discussed widely—in public policy forums, the medical and scientific literature, and the lay press. This is a symptom of an important change in health care markets. Attention has shifted from the "passive" funding and administration of systems to active concern about the cost of care and the health outcomes achieved. The health economic thinking that now permeates health policy and health care systems is raising questions such as: How much should we spend on health care and how do we ensure it is spent efficiently? How and when should we assess the outcome of using health technologies in clinical practice to ensure resources are used efficiently?

Box 1.1 Definition of health economics

Health economics is the application of the theories, tools and concepts of the discipline of economics to the topics of health and health care.

Economics as a science is concerned with the allocation of scarce resources; health economics is concerned with the allocation of scarce resources to improve health. This includes both resource allocation within the economy to the health care system and within the health care system to different activities and individuals.

A range of approaches to economic evaluation has been developed to help address these important questions of efficiency. This guide provides an introduction to them. The first chapter reviews contextual background, illustrating the increased level of interest in the use of economics by policy makers, payers, and health care providers. Chapter 2 introduces the various types of economic evaluation and discusses how they approach the two components of economic evaluation: what effect a treatment has on health and what it costs. The challenges are illustrated with examples of cost-of-illness studies, which seek to quantify the aggregate costs of a disease and its treatment. Chapter 3 explores the methods of economic evaluation in greater detail, focusing particularly on the use of modelling techniques that synthesise data from a range of sources. The chapter illustrates these techniques using a number of

Box 1.2 Definition of health technology assessment

Health technology assessment (HTA) is a multidisciplinary process that summarises information about the medical, social, economic and ethical issues related to the use of a health technology throughout its life span in a systematic, transparent, unbiased and robust manner.

The aim of HTA is to inform the formulation of safe and effective health policies that are patient focused and seek to achieve value for money.

examples, primarily from evaluations of drugs. Important aspects of each methodology are explained and particular challenges identified. Chapter 4 discusses methodological guidelines for the conduct of the economic evaluations that are required or suggested in several countries. Chapter 5 concludes.

Challenges in health care: the context

Total health care spending as a proportion of gross domestic product (GDP) has steadily increased in all OECD countries, albeit starting from different levels. Spending in the European Union was between 7.5% and 12% of GDP in 2010 (9.5% to 12% in Western Europe, 7.5% to 9.5% in Central/Eastern Europe). In the US, it reached over 17% of GDP (see Table 1.1).

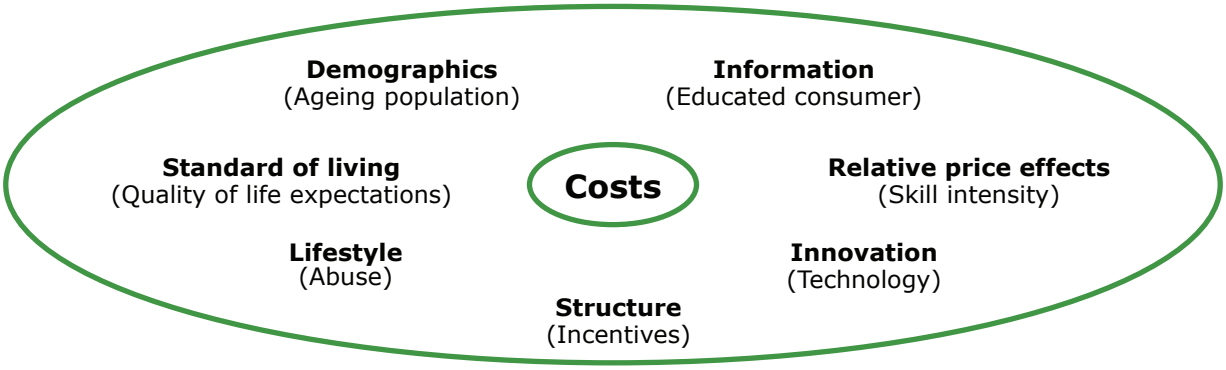
Table 1.1. Health care expenditures as percentage of GDP

Country or Region	1970	1980	1990	2000	2010
OECD average	5.8%	7.3%	8.7%	8.4%	9.3%
US	6.9%	8.7%	11.9%	13.2%	17.7%
Japan	4.6%	6.5%	6.1%	7.7%	9.6%
Western Europe	4.7%	6.7%	7.1%	8.4%	9.5%
Central/Eastern Europe	NA	NA	NA	6.1%	7.9%
Latin America	NA	NA	NA	NA	7.0%

Source: OECD (2013), WHO (2011)

Numerous interdependent factors contribute to increased health care costs, as indicated in Figure 1.1. In the industrialised world, the elderly population often is singled out for concern as it consumes a substantial and increasing share of health care resources. Health care expenditures have risen less because of demographic change, however, than because of the availability of a greater number of treatment options and continuous improvement in the quality and intensity of care. More can be done, so more is done.

Figure 1.1. Major contributors to the growth of health care costs



Concerns about the financing of health care are high on every government's agenda, particularly in countries where health care is predominantly funded with public money via taxes, social insurance or a combination of the two (see Table 1.2). Among the OECD countries, the US is an exception, with most health care being financed by private insurance, although public financing is increasing steadily. The private portion of the market in Latin American countries is substantial and growing.

Governments around the world, and particularly in Europe, have attempted to contain costs using a variety of measures aimed at both the demand for and the supply of health care. Figure 1.2 shows those that have been aimed at the prescription pharmaceutical market. These measures, however, have been less successful than hoped, partly because growth in spending is driven primarily by the availability of new and improved technology to which cost-containment measures are less easily applied.

Table 1.2. Public health expenditures as percent of total health expenditures

Country or region	1970	1980	1990	2000	2010
OECD average	73%	73%	73%	72%	72%
US	36%	41%	39%	43%	48%
Japan	70%	71%	78%	81%	81%*
Western Europe	76%	76%	77%	76%	76%
Central/Eastern Europe	NA	NA	NA	75%	72%
Latin America	NA	NA	NA	NA	52%

* Data for 2009
Source: OECD (2012), WHO (2010)

Figure 1.2. Examples of measures for containing spending on prescription drugs in Europe

Pricing	Clustering (same price for similar treatments) Price cuts, price freezes Reference pricing
Listing	De-listing (removal from eligibility for reimbursement) Positive or negative lists of products eligible for coverage
Shaping use	Greater use of generics and/or control of generic prices Increased patient co-payment Prescribing budgets and/or guidelines for doctors
Purchasing	Tendering Volume contracts
Indirect cost control	Profit limits for manufacturers Promotional budget limits for manufacturers Reductions in wholesale and retail pharmacy margins

The financial crisis that began in the late 2000s has exacerbated the situation by making further increases in public spending on health care more difficult. Discussions and decisions about prices and purchasing, as a result, are now taking place in an environment characterised more by concern about cost and value than about demand for innovation. Health care decision makers everywhere are focusing more narrowly on efficiency and within tighter budgets. New, more expensive, therapies must carry a clear additional health benefit to be deemed worthy of an additional expenditure. Decisions makers, then, will increasingly require that innovative therapies—medicines and other interventions—be assessed for relative effectiveness and cost-effectiveness, rather than only efficacy and safety (see Figure 1.3).

Figure 1.3. Assessment criteria for new therapies

Safety	Does it have side effects and are these acceptable and manageable?
Efficacy	Does it work in a controlled environment (clinical trials)?
Relative efficacy	How well does it work in a controlled environment compared to one or more alternatives (standard treatment)?
Effectiveness	Does it work in normal clinical practice?
Relative effectiveness	How well does it work in normal clinical practice compared to other alternatives (standard treatment)?
Cost effectiveness	Is it an efficient use of resources, i.e. is an additional benefit worth an additional cost?

A number of European countries long have requested cost-effectiveness assessments as an aid in deciding about the reimbursement status or price of a new technology. Demand is growing for comparative trials that can better define the incremental benefit of a new treatment. Cost-effectiveness, and even comparative analyses, however, are based on models created before the product reaches the market. Until a product has been used in routine clinical practice, considerable uncertainty remains about both clinical outcome and resource use. As a result, authorities increasingly are requesting additional evaluations using experience from actual clinical practice. In some cases, the results can lead to a renegotiation of the price and also may be used to shape clinical practice.

The Role of Health Economic Evaluation Studies in Market Access

An economic evaluation is a tool for assessing the benefits and costs of competing uses of scarce resources. It provides data in a structured format that is comparable across diseases, but does not in itself offer a decision. Since value for money is now a core concern, analyses of the consequences of the use of new and existing therapies, in terms of both benefits and costs, have become essential to decisions about resource allocation. Cost-effectiveness has become an important criterion not only for deciding which therapies ought to be funded or reimbursed, but also for identifying the patient populations that should have access.

Figure 1.4. Definition and forms of economic evaluation

Definition of economic evaluation: A comparative analysis of two or more options in terms of their costs and consequences

Types of economic evaluation	
Cost-minimisation analysis (CMA)	Comparison of costs of alternatives that have the same health outcome Allows comparison within a clinical indication
Cost-effectiveness analysis (CEA)	Comparison of costs and disease-specific health outcomes (e.g. life-years saved, patients cured, events avoided) Allows comparison within a clinical indication
Cost-utility analysis (CUA)	Comparison of costs and generic health outcomes (e.g. quality-adjusted life years) Allows comparison across clinical indications
Cost-benefit analysis (CBA)	Comparison of costs and health outcomes valued in monetary terms (e.g. willingness to pay) Allows comparison to other sectors of the economy

Many countries have official or quasi-official specialised groups that assess the value of both current and new health care technologies. These may be independent reimbursement agencies or specialised HTA agencies. Economic evaluations are an integral part of their assessments.

An economic evaluation provides a comparative analysis of alternative courses of action in terms of costs and consequences (see Figure 1.4). This entails comparing alternative treatment strategies over the entire course of a disease, or defined disease episode, in order to identify the best option for specific patient groups, given expected costs. Such evaluations use aggregate measurements and provide information for groups of patients, rather than individual patients. All evaluations use similar techniques to estimate cost, although different techniques are used for measuring consequences, depending on the disease or the desired result.

When two interventions have the same outcome, the less costly one dominates and is preferred. Interest is greater in products that improve outcomes compared to existing treatments that are only equivalent in outcome. But more efficacious technologies generally come at a higher cost. Thus, an

incremental cost-effectiveness ratio (ICER), i.e. the extra investment required for the additional health benefit, is computed. The more costly intervention will be adopted if the incremental cost per unit of health effect is less than the purchaser's willingness to pay for such a health gain.

Box 1.3. Definition of an ICER

$$\frac{[\text{Cost (B)} - \text{Cost (A)}]}{[\text{Effect (B)} - \text{Effect (A)}]} \quad \text{or} \quad \frac{\text{Difference in Cost}}{\text{Difference in Effect}}$$

where B is more effective and more expensive than A

(if B is more effective and less expensive than A, B dominates A and the ICER is not calculated)

Health-related costs may be incurred in a range of social spheres, making it important to include all costs for a relevant time period, even if they fall under different budgets. For instance, a new treatment may increase the pharmaceutical budget, but over time produce enough savings in other parts of the system to partly or fully offset this increase, such as lower hospitalisation costs or fewer monitoring requirements. Savings also may occur in other sectors of the economy, for example, when sickness absences, early retirement due to disease, or premature deaths are avoided. For efficient resource allocation, decisions should consider the full impact of therapies, regardless of where effects occur. Economic evaluations, then, must start from a societal perspective to capture all potential benefits.

Adopting a societal perspective to assessing the value of treatments matters and makes sense (Jönsson, 2009; Johannesson et al, 2009). Regulatory authorities take a societal perspective in licensing a drug, weighing risks against wider benefits. Economic analyses, similarly, need to include both costs and benefits to society overall. This can help decision makers avoid an overly narrow, budget-specific perspective, which may miss the important benefits accrued outside that budget and produce suboptimal decisions about resource allocation. Narrow decisions may inappropriately restrict access by not funding the treatment at all or by inappropriately limiting it to only some groups of patients. In such cases, the payer may achieve the objective of controlling the budget (static efficiency), but the greater benefit to society, particular patients, will be missed (dynamic efficiency).

Despite the rather obvious potential benefit of using health economic evaluations, decision makers across Europe vary in how and how much they are used. The remits of decision making organisations also differ: HTA agencies are generally concerned with whether or not to recommend treatments, but lack the power to decide on access and price; some reimbursement agencies can only accept or refuse to fund a treatment at the proposed price, while others have the power to negotiate price (Figure 1.5).

In recent years, HTA agencies have become increasingly involved in decisions about early market access, blurring the distinction between their activities and those of traditional reimbursement assessments. For example, the National Institute for Health and Care Excellence (NICE) within the National Health Service (NHS) in the UK assesses selected new treatments early on and its recommendations are binding. Decisions by the Scottish Medicines Consortium, an HTA body within the Scottish NHS, are fully binding. In France, the Haute Autorité de Santé (HAS) includes bodies that assess the absolute and relative benefit of a new technology and its reimbursement status, and those that perform full assessments of technologies after they have entered the market. In Germany, the Institute for Quality and Efficiency in Health Care (IQWiG) assesses the effectiveness—and, if requested, cost-effectiveness—of new treatments one year after their introduction. Clearly, then, the timing and impact of cost-effectiveness studies varies across countries and organisations.

Guidelines for performing economic evaluations have been produced in many countries. These fall into two categories.

1. Reimbursement guidelines, i.e. guidelines issued by authorities that make the submission of economic evaluations mandatory for listing a new product on the reimbursement formulary, and that define the format of such submissions

2. Methodological guidelines, i.e. guidelines proposed by researchers or groups of researchers with the aim of improving the techniques and methods used and making studies more transparent

Figure 1.5. Bodies involved in determining market access

	Regulatory agencies	Reimbursement agencies	HTA agencies
Role	Market authorisation; subsequent review of benefit-risk profile, if warranted	Coverage decision within a health care system, given resource constraints	Provide best evidence to inform coverage decisions (e.g. clinical practice guidelines)
Evidence used	At launch: safety and efficacy (potentially relative efficacy) data from randomised clinical trials Post launch: safety follow-up	At launch: relative efficacy/effectiveness and budget impact, formal cost-effectiveness analyses in most countries Post launch: relative effectiveness and cost-effectiveness	At launch: seldom involved Post launch: relative efficacy/effectiveness, cost-effectiveness
Power	Decision	Decision (with/without price negotiation)	Recommendation

The first country to make submission of economic studies an official requirement for listing medicines on the national drug formulary for reimbursement was Australia, in 1993. Since then, the guidelines for submissions to Australia's Pharmaceutical Benefits Advisory Committee (PBAC) have been updated several times to incorporate experience gained (PBAC, 2008).

The second country to require economic studies was Canada, based on an initiative in the province of Ontario. Detailed methodological guidelines were developed in collaboration with all stakeholders: government, insurance companies, providers' associations (hospitals, pharmacists, physicians), academia and the pharmaceutical industry. Revised editions were published in 1997 and 2006, with addenda in 2009 that covered indirect treatment comparisons and evaluations in oncology. The Canadian document is widely considered authoritative in terms of methodological standards and most of the guidelines published subsequently by other agencies have relied heavily on the Canadian guidelines (CADTH, 2006 and 2009).

Initially, European countries took a somewhat different approach. While guidelines as an expression of methodological standards were elaborated and published in most countries, they were not at first tied to reimbursement decisions. Now, however, the majority of countries have made economic evaluations mandatory for reimbursement decisions and require studies to follow official guidelines produced by the reimbursement authorities. (See Figure 1.6 for a non-exhaustive list of guidelines.) Differences among the guidelines are limited, with the most important being the perspective that submissions are expected to adopt. Other differences relate to discount rate, time horizon, and level of detail in forecasting use of a new product, i.e. the anticipated budget impact. As many of the countries that have made these studies mandatory are rather small, they minimize additional effort by accepting the results of studies from other countries, with appropriate adaptation to local needs.

In the US, the Department of Health and Human Services commissioned a panel of academic experts, the "Washington Panel", to elaborate a set of guidelines for good practice. The effort produced a widely-quoted book (Gold et al, 1996) that has sparked intense scientific discussion aimed at further development of the methods. Since then, the Academy of Managed Care Pharmacy has published a more specific set of guidelines for submissions: *AMCP guidance for submission of clinical and economic*

Figure 1.6. Use of economic evaluation in various countries

Country	Use of economic evaluation	Formal Guidelines Year of 1st publication	Research Guidelines Year of 1st publication
Australia	Required for all new drugs	1993	NA
Austria	Required for all outpatient drugs, with focus on comparison budget impact and price	NA	2006
Belgium	Required for all outpatient drugs, with focus on added benefit assessment	NA	2002
Canada	Required at national and provincial level	1995	NA
Denmark	Voluntary submission	NA	1997
Finland	Required for all outpatient drugs	1999	NA
France	Reimbursement only based on added benefit; re-assessment by HTA agency	2011	2004
Germany	Upon request, one year after launch	2010	1995
Hungary	Required for all drugs	2002	NA
Italy	Authority to request at national and regional level	NA	2001
Netherlands	Required for all new drugs outside existing clusters	1999	NA
New Zealand	Required for all new drugs	1993	NA
Norway	Required for all prescription drugs	2002	NA
Poland	Required for innovative drugs	2007	NA
Portugal	Required for all new outpatient and inpatient drugs	1999	NA
Spain	Not required at national level; can be used at regional level	2010	1995
Sweden	Required for all new drugs	2003	NA
UK (England & Wales)	Submissions requested on defined drugs and devices, either for review of class or for single technology appraisal	1999	NA
UK (Scotland)	Required for all new drugs and devices	2000	NA
USA	Inconsistently used for listing	NA	1996; 2001

evaluation data to support formulary listing in US health plans and pharmacy benefits management organisations (Sullivan et al, 2001).

In addition to the documents and guidelines produced by individual countries, a group of academic researchers published a report on researcher independence in 1995 that attempts to deal with problems of bias in economic evaluation (Task Force, 1995). The report suggests that evaluations ought only to be performed by independent researchers with no direct financial link to the sponsor or, if a study is sponsored, researchers should have complete freedom to publish any and all results. This is based in part on concerns about inappropriate modification, at a later stage, of elements such as effectiveness measures and analytical methods. Unlike protocols for clinical trials, those for the economic evaluation of new drugs are not always defined in detail at the outset. However, the solution to potential ethical problems such as this surely must lie in adherence to good practices by all participants in this evolving field, rather than in contractual arrangements.

Similarly, allegations that only studies with positive results are published indicate a fundamental misunderstanding of the purpose of economic evaluation. First and foremost, economic evaluation studies are a tool to support decisions about resource allocation. The primary purpose of such studies, then, is not to achieve publication, but to inform decision making. By nature, they are not hypothesis testing in the way that clinical trials are, but instead seek scenarios where the product under evaluation can be expected to be cost effective. The scenarios may involve specific patient populations (subgroups), specific administrative conditions, specific positioning (first-line or second-line therapy, last resort), and so on—all variables that are informed by the clinical trial results and hence often cannot be specified in a general set of guidelines beforehand. The goal of payers is to make treatments available in an efficient way, i.e. to those patients most in need and in those settings where they are cost effective. "Negative" results (i.e. high ICERs) are thus of no interest except for rejecting that particular scenario. The only way to ensure both credibility of the claims of value for money and usefulness of the studies to decision makers is to use sound methodology and relevant data, and to report results in a complete and transparent manner.

Among those countries where economic analysis must be considered prior to deciding on reimbursement for new products, economic submissions also are required when approval is sought for a new indication for an existing treatment. But as is apparent from Figure 1.5, considerable differences exist in the extent to which economic analysis is used. Sweden and Finland informally used economic evaluations in decision making even prior to the systematic assessment of all new technologies. In The Netherlands, an economic criterion is applied to reimbursement decisions only for drugs that cannot be included in an existing therapeutic cluster under the reference pricing scheme. In Norway, all new products for general prescription (schedule 2) require an economic submission, while in Portugal both outpatient and hospital drugs are subject to economic evaluation. Belgium and Austria both require economic evaluations, but Belgium appears to have a strong focus on added benefit while Austria appears to focus on price comparisons and budget impact. In Scotland, funding decisions based on cost-effectiveness, among other parameters, are binding.

Among the large countries in Western Europe, only the UK has truly formalised its requirements. NICE was set up in 1999 by the Department of Health to assess new and existing health technologies and recommend whether and how these technologies should be used within the NHS in England and Wales. Since the beginning of 2002, in an effort to limit regional differences in access, it has been obligatory for the NHS to fund prescriptions based on NICE's recommendations.

The organisation and functioning of NICE are different from similar agencies in other countries, in part due to the long tradition of academic research in health economics in the UK, coupled with a drive for greater transparency and public discussion.

> *NICE's role is to improve outcomes for people using the NHS and other public health and social care services by:*
>
> - *Producing evidence-based guidance and advice for health, public health and social care practitioners;*
> - *Developing quality standards and performance metrics for those providing and commissioning health, public health and social care services;*
> - *Providing a range of information services for commissioners, practitioners and managers across the spectrum of health and social care* (NICE, 2013b).

A sizeable component of NICE's work has been made up of technology appraisals, which may examine complete indications, single technologies, or entire classes of drugs.

- Is the technology likely to result in a significant health benefit, taken across the NHS as a whole, if given to all patients for whom it is indicated?
- Is the technology likely to result in a significant impact on other health-related Government policies (for example, reduction in health inequalities)?

- Is the technology likely to have a significant impact on NHS resources (financial or other) if given to all patients for whom it is indicated?

- Is there significant inappropriate variation in the use of the technology across the country?

- Is the Institute likely to be able to add value by issuing national guidance? For example, in the absence of such guidance is there likely to be significant controversy over the interpretation or significance of the available evidence on clinical and cost effectiveness? (NICE, 2013c)

When developing technology appraisals guidance, NICE commissions an independent academic centre to review the existing published evidence on each technology and, in some cases, the evidence contained in the manufacturer's submission. It also may ask the academic group to perform an independent economic evaluation. A specific guidance for manufacturers has been developed to ensure that all submissions have the same format (the "reference case").

The Importance of Economic Evaluation for the Development of New Technologies

Economic evaluations have become a key element, and in many countries, a mandatory requirement, in supporting reimbursement submissions. In most pharmaceutical companies, these studies are an integral part of research portfolio management intended to developed products for the market that the market wants. A similar development is underway in the medical devices industry. However, once reimbursement status has been achieved, little attention has been given to ensuring that products still offer value for money when used in actual clinical practice. This is changing gradually and an increasing number of countries now review reimbursement decisions at regular intervals (e.g. Canada, France) or periodically (e.g. Sweden, England/Wales).

Reassessment may be a particular challenge in some cases—for instance, in chronic diseases where the treatment goal is to delay progression to severe disease states with high costs and low quality of life (e.g. multiple sclerosis, rheumatoid arthritis), or in disease areas where most treatments aim to prevent mortality (e.g. heart disease, cancer). For such diseases, it may take a number of years before it is possible to observe the effect of a new treatment in the "real world".

Economic evaluations at launch are by definition based mostly on relative efficacy observed on the controlled environment of the clinical trial. Assessing relative effectiveness and hence cost-effectiveness in the "uncontrolled" clinical practice environment presents quite different challenges. However, conditional reimbursement approvals are becoming common, tying initial reimbursement to subsequent proof of cost effectiveness in clinical practice. Contractual agreements where the financial risk is shared between manufacturers and the health care system also require economic evaluation based on clinical use. Perhaps the greatest challenge presented is availability of relevant cost and outcome data from clinical practice. Observational follow-up, cohort studies and patient registries can supply such data, provided they are set up do so.

The demand for comparative data already exerts a substantial impact on the clinical development of new treatments, a situation that is likely to intensify. Marketing authorisation traditionally has been based only on efficacy and safety evidence for the particular product. The current demand for improved, rather than similar, outcomes, however, requires comparative studies that consider relative efficacy. The choice of comparator can have crucial implications: in addition to the difficulty of choosing a comparator that is deemed an appropriate alternative treatment option in the largest number of markets, the choice made also may drive the positioning and/or the price of the new product.

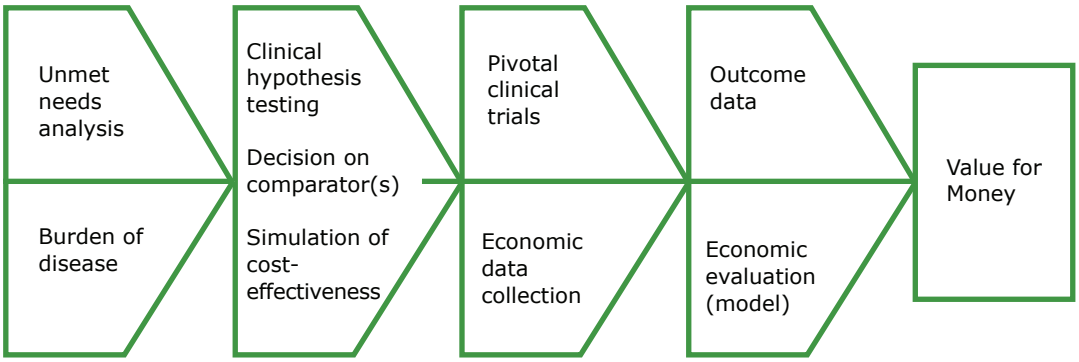
While requirements for comparisons are being better defined, reimbursement authorities and HTA agencies usually accept indirect comparisons between treatments. To give a simple example, if treatment A has been compared to treatment B, and treatment C has also been compared to treatment B, it is possible to statistically estimate the comparison of A and C. This clearly is less certain than direct comparison, particularly if the studies of B versus C were performed some years prior to those of A versus B (or vice versa). Despite this limitation, agencies tend to take the pragmatic view that it is better to have at least some supporting evidence available for decision making.

Integrating comparative research into the development process and combining clinical and economic objectives presents a number of challenges.

- How can efficacy be translated into effectiveness?
- What is an appropriate outcome measure? How can a patient's health outcome be transformed into a quantifiable measure, e.g. quality of life (utility)?
- What is an appropriate time frame for such economic analysis, compared to clinical proof of efficacy?
- What is the appropriate product or other intervention for comparison? And how can comparisons against placebo be incorporated?
- Where and how can resource-use data be collected?

Some of these points are addressed in methodological guidelines. More often, however, the chosen approach is guided by feasibility based on time frame, resource constraints, data availability, and the indication and positioning of the new treatment. Clinical trials generally are regarded as inadequate vehicles for collecting data on resource because consumption in a trial is mandated and heavily influenced by the protocol. An illustration of the overall combination of clinical and economic development is provided in figures 1.7 and 1.8. Figure 1.7 shows how the accumulation of information produces evidence of value for money; Figure 1.8 provides details about the sequence and phase timing of economic evaluation.

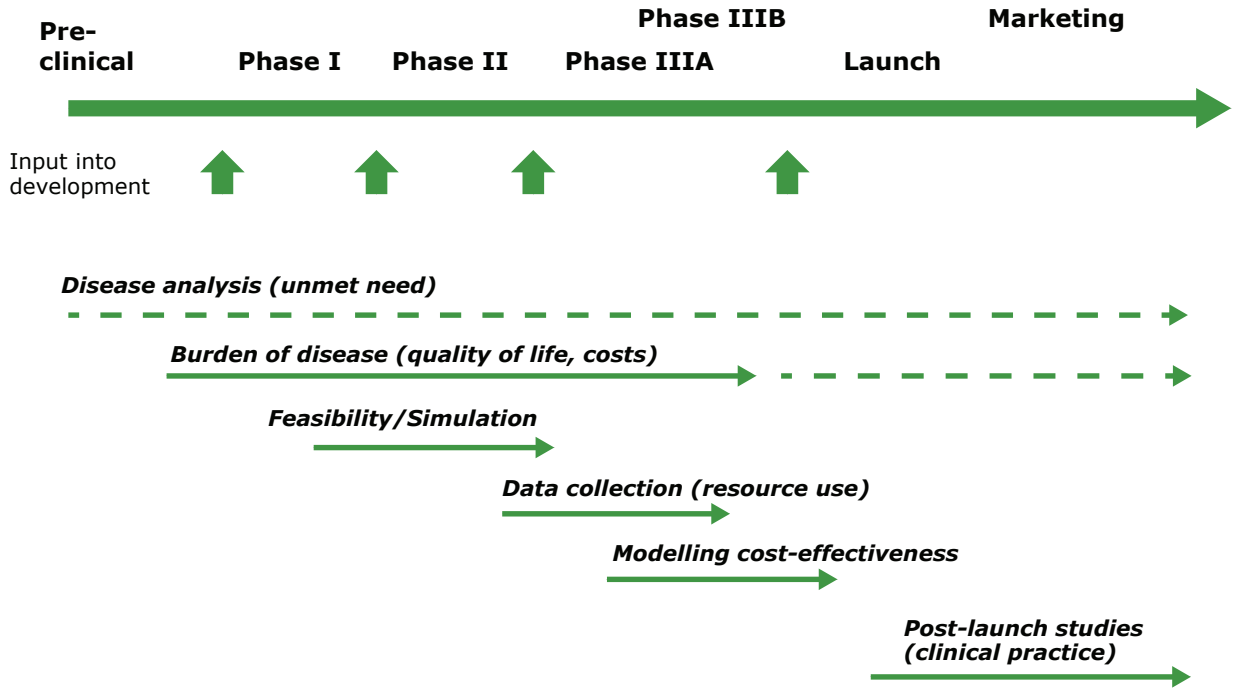
Figure 1.7. Documenting value for money



The evaluation process spans the entire development time for new products. It will be more successful if performed with due regard to the anticipated information needs of providers and payers, and if fully integrated into the clinical development process. In the earlier stages of development, activities mostly involve basic research about the disease, its economic consequences and the costs of treatments. In later stages, economic data are collected while Phase III clinical trials are taking place. Because economic evaluations typically consider a wider frame and longer time horizon than clinical trials, data from different sources may need to be combined: data on the disease and its development (epidemiological data), data on patient management and resource consumption (economic data), and outcomes data (clinical trials, registries). Most economic evaluations thus are modelling studies by default and such studies are now accepted as the rule, rather than the exception, by reimbursement authorities and HTA agencies.

Conducting extensive economic evaluation at all stages of development can be expensive and the knowledge gained limited, both because of the nature of clinical trials and because data about the most effective use of a product accumulates only over time in actual use. Studies conducted after launch certainly are not without cost; incentives for such expenditure are only now developing. A balance must be struck between the costs and the benefits of preparing economic evaluations throughout a product's life cycle. Generating economic information that will not be used or that could be misleading is pointless.

Figure 1.8. Economic evaluation during and after development



Chapter 2

FORMS OF HEALTH ECONOMIC EVALUATION

Introduction

A health economic evaluation is a means of establishing the "value for money" of health care technology and as such is an integral part of HTA. Taking as our starting point the definition of an economic evaluation in health care as "a comparative analysis of alternative courses of action in terms of both their costs and consequences" (Drummond et al, 2005), economic analyses are always comparative and are applied to explicit alternatives. One pharmaceutical product can be compared more or less with itself (different dosages or modes of administration), or with another pharmaceutical product, or with another type of intervention such as surgery, or with a "watchful waiting" approach whereby the patient receives no form of medical intervention, but instead is monitored for any change in health status. A treatment cannot be cost-effective by itself, but only in relation to one or more relevant alternatives and for defined patient groups. Whatever the alternative, at a minimum all the costs related to each method of treating a relevant disease episode must be considered and related to the benefits in terms of improvement in the length or quality of life.

All forms of economic evaluation involve assessment of both the inputs (the use or loss of resources) and outputs (health benefits) of the health care programmes to be compared so as to facilitate the process of choosing the most appropriate alternative. The decision criterion is to maximise health outcome for the population as a whole (social utility), given resource constraints. If a treatment strategy generates better outcomes and is less costly, it dominates the alternatives. More often, however, a treatment strategy that generates better outcomes also will be more expensive and, as was noted earlier, a judgement will have to be made as to whether the incremental benefit is worth the incremental cost.

Box 2.1 illustrates the structure of economic evaluation. The inputs, or costs, are defined as the costs related to the use of the treatment minus the costs that are avoided as a result of its use, compared to costs without the treatment or with a different treatment. Costs are a function of the quantity of resources used and their price. Detailed data on prevailing treatment strategies in clinical practice, however, are seldom readily available. Identifying the relevant resources, quantifying and valuing them, then, is generally necessary, but it is a rather straightforward process. Outputs are more difficult to estimate for several reasons. Treatments often affect multiple symptoms or events and at different points in time. It may not be obvious how to combine these effects into a single comprehensive outcome measure.

Box 2.1. Components of economic evaluation

INPUTS defined as resources used or lost.

Direct costs are defined as costs related to the use of resources due to either the disease or its treatment. We generally distinguish between

- Costs to the health care system *(direct medical costs)* and
- Costs to social services and to patients themselves or to their relatives *(direct non-medical costs).*

Indirect costs or loss of production are defined as costs that occur to society related to loss of production, due either to the disease or its treatment. We generally distinguish among

- Short term losses due to sickness absence
- Long term losses due to premature death or early retirement due to the disease (invalidity)
- Losses due to reduced productivity while at work due to the disease (e.g. because of fatigue, migraine attacks).

Figure 2.1. Structure of economic evaluation

Other costs that occur due to illness and may be influenced by treatment are *intangible costs.* These relate to the suffering and loss of quality of life experienced by the patient, and sometimes are included in descriptive studies (cost-of-illness studies). In the framework of a cost-utility analysis, the effects of a treatment on quality of life are included in the health outputs (as part of quality-adjusted life years [QALYs]). Intangible costs are particularly difficult to measure and value. Several approaches exist, including the use of quality of life instruments, direct measurements within the framework of willingness to pay assessments, or a valuation where the loss of QALYs compared to the normal population is valued with an assumed willingness to pay for making up this loss.

OUTCOMES are measured as health improvements expressed as

1. Disease measures such as events avoided or delayed (e.g. hip fractures in osteoporosis; myocardial infarction, stroke or death in cardiology), patients successfully treated (e.g. number of cancer patients in complete remission; number of infections cured within a given time)
2. Survival measured in terms of lives saved or life-years saved
3. Quality-adjusted survival, expressed as QALYs
4. Monetary value, expressed as willingness to pay for the improvement

Types of Economic Evaluation

Economic evaluations are categorised by type, distinguished primarily by how outcomes are treated. The appropriate means of evaluating outcomes will depend on a number of factors, the most important being the medical and economic problem addressed—i.e. whether the evaluation seeks to inform the selection of a treatment for patients with the same disease, or to inform the prioritisation of treatments for different diseases. The medical question will determine what effectiveness measure is used, while the economic question will influence both the effectiveness measure and the type of evaluation to be used. In general:

1. If the economic question is whether a treatment is a good use of resources within the disease area, the comparison is with similar treatments and the outcome measure can be disease specific. The type of evaluation will be a cost-effectiveness analysis if there is only a single outcome. With multiple outcomes, it is necessary to choose one, or to construct an index. For example, outcomes in hypertension can be stroke or chronic heart disease; in

osteoporosis, several different types of fractures can happen; and in cancer, outcomes can be measured in terms of survival, remissions, side-effects, quality of life, etc.

2. If the economic question is whether a treatment represents a good investment considering the entire spectrum of diseases, the comparison will be with treatments for other diseases and the outcome measure will need to be generic, such as the QALY, which is a combination of life expectancy and quality of life. This will enable a cost-utility analysis, a specific type of cost-effectiveness analysis. It is appropriate to conduct a cost-utility analysis when quality of life is an important component of the effect of the disease and its treatment, or when there are a large number of different symptoms and effects to consider.

A somewhat different form of economic analysis is the cost-of-illness study, described in detail later on. A cost-of-illness study is not evaluative, but purely descriptive; it aims to establish and quantify the burden that a particular disease places on society. Since these studies do not consider the outcome of treatment, they are of limited value to decision makers concerned with achieving value for money in health care. However, they provide important background information on the disease and its cost—an overall economic assessment of the current situation. As such, they can provide much of the basic data for an economic evaluation that investigates the outcome when something in that situation changes, such as a new treatment being introduced.

By far the most important question to ask before embarking on an economic evaluation is whether or not clear and well-documented clinical evidence is available for the technology to be compared to the available alternative(s). An economic evaluation can only be as good as the underlying effectiveness data, and the highest quality economic data will not be able to overcome any deficiency in the effectiveness data. Data quality has become one of the most important topics in the current debate surrounding HTA where the demand is now for comparative effectiveness data using patient-relevant outcome measures, rather than data on efficacy against placebo.

Figure 2.2 summarises the effectiveness measures used in the different types of evaluation and indicates what questions each type of evaluation typically addresses. Each of these analyses is discussed in detail later and illustrated using examples.

Figure 2.2. Effectiveness measures used in economic analyses

Type of analysis	Effectiveness measure	Decision support
Cost-minimisation analysis	No measurement, as cost-minimisation analysis applies to alternatives with no difference in outcome	Comparison of treatments for the same disease
Cost-effectiveness analysis	One disease-specific patient-relevant measure, such as events avoided (e.g. stroke, relapses), cure, disease-free time, or a more general measure such as life-years saved	Comparison of treatments for the same disease
Cost-utility analysis	A summary measure combining survival and quality of life (i.e. quality-adjusted life years)	Comparison of treatments for different diseases
Cost-benefit analysis	Effectiveness expressed as a monetary benefit (e.g. willingness to pay for a given effect)	Comparison of investments in the health sectors with those in other sectors (e.g. education, road safety)

Outcome Measurement in Economic Evaluation

In clinical trials, as in clinical practice, several different measures can be used to express health outcomes because a variety of treatment effects may be important in terms of clinical management. In

economic evaluation, on the other hand, outcomes need to be expressed using a single effectiveness measure that is easy to understand and to relate to the disease, and that ultimately can be compared to outcomes across diseases. The measure should also express the overall and final outcome, rather than intermediate ones.

In acute and curable diseases, such as infections, it is rather straightforward to define the final outcome in a dichotomous way, such as "cure" or "no cure". The economic evaluation then will estimate and compare the costs of achieving the cure using different treatment strategies. For example, if a new treatment cures an additional 10% of patients than is currently the case at an additional cost of €100 per patient, then the cost per extra cure achieved is €1000 (€100/0.1).

In disease areas where the risk of an undesirable event is continuous, such as cardiac disease, the outcome may be defined as avoiding or postponing that event. However, the ultimate objective of preventing serious clinical events is to avoid the consequences of the event (such as death or serious disability), rather than the event itself. Economic evaluation thus will preferably attempt to capture the consequences of avoiding such clinical events by estimating changes in survival and quality of life.

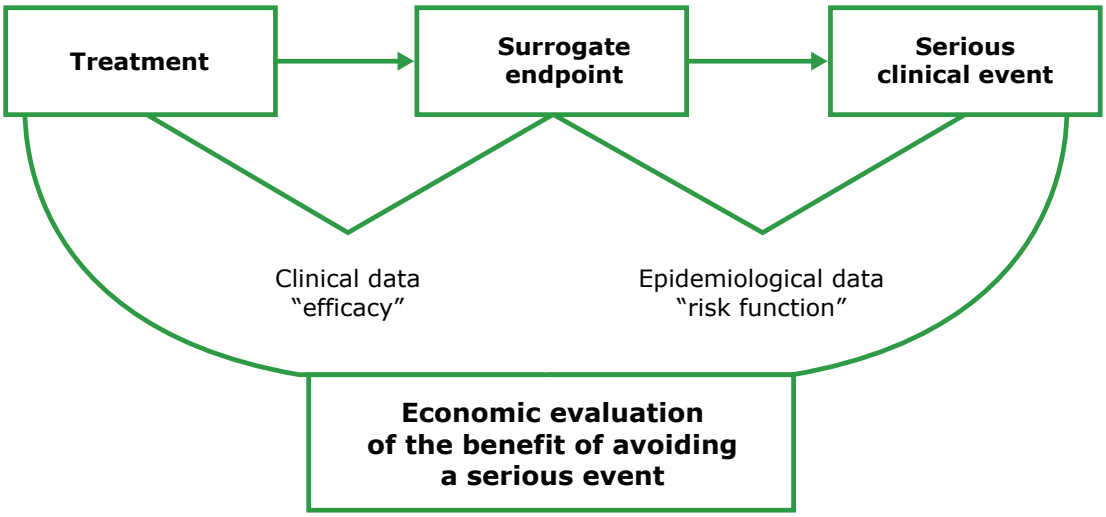
In chronic diseases, on the other hand, particularly in chronic progressive diseases, defining an overall final outcome is more difficult and efficacy is often assessed based on intermediate endpoints only. Some of these endpoints are patient related, even if they do not express the final outcome, such as exacerbations, relapses and recurrences of the disease. Some of the endpoints assessed in clinical trials are inadequate for translation into effectiveness, such as a relative improvement of 20%, 50% or 70% in a grouping of multiple symptoms—as in rheumatic diseases, for example.

Physiological measures and clinical events

Physiological measures (or surrogate endpoints) such as mmHg in hypertension, mMol cholesterol in hyperlipidaemia, or bone mineral density in osteoporosis, are routinely used in clinical management as outcome measures as they are linked to clinical events such as stroke, myocardial infarction and fractures. In these cases, economic evaluation can then estimate the value of avoiding (or postponing) an event, provided that epidemiological data linking the surrogate measure to the undesirable event are available. The cost-effectiveness of treatment today which aims to avoid a future event can be estimated if it is possible to derive a risk function for the annual risk—for example, of a hip fracture at a given level of bone mineral density and at a given age, or of a myocardial infarction at a given level of cholesterol, controlled for age, gender and other known risk factors (such as smoking).

Figure 2.3 illustrates this concept. Here, a risk function for the annual risk of a serious clinical event at given levels of a surrogate measure and under different conditions (age, sex, risk factors) is derived from epidemiological data. This links short-term intermediate endpoints with final outcomes, enabling a calculation of the cost-effectiveness of treatments that reduce this risk.

Figure 2.3. Extrapolating from intermediate to final outcome



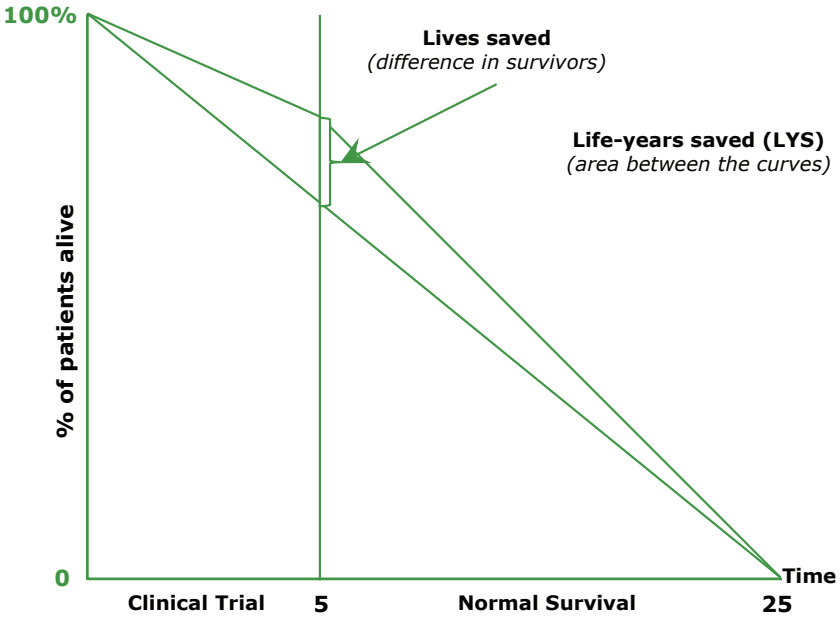
Survival

As noted above, many events may not represent the true final outcome in a disease as they may be important only insofar as they are linked to mortality risk. In diseases without clearly defined events, survival is often the relevant endpoint.

Survival can be expressed in different ways, for example: the proportion of patients alive in each group at the end of a clinical trial, the number of deaths avoided, the number of patients alive after five years, or overall survival. In economic evaluation, survival is generally measured in terms of years of life, and is represented by an area under the survival curve that can be related to both costs and quality of life. However, clinical trials are seldom long enough to provide the data necessary to estimate directly the number of life-years saved (LYS) by one treatment compared to another. Epidemiological data are again required to extrapolate from the short-term perspective of lives saved to the long-term perspective of life expectancy.

Figure 2.4 illustrates the concept of LYS and shows that the effects of a treatment achieved within trials carry over to the period after the trial. A difference in the number of patients alive at the end of a clinical trial will lead to a difference during the years after the trial. For instance, if we assume that 5% of patients surviving at year five die every year after the end of the trial in both the control and intervention groups, all patients will be dead after 20 years. Mean and median survival after the trial will be ten years. If we further assume that survival at the end of the trial was 80% in the control group and 90% in the intervention group, the gain in life expectancy in the intervention group will be 0.25 years during the trial ((5×0.1)/2) while the gain after the trial will be one year ((20×0.1)/2). The life expectancy at the start of the trial will be 12.5 years in the control group and 13.75 years in the intervention group, with the majority of the difference achieved after the clinical trial. The area between the two curves in the Figure 2.4 represents the difference in life expectancy of the two groups.

Figure 2.4. Extrapolating from within-trial mortality to life years saved



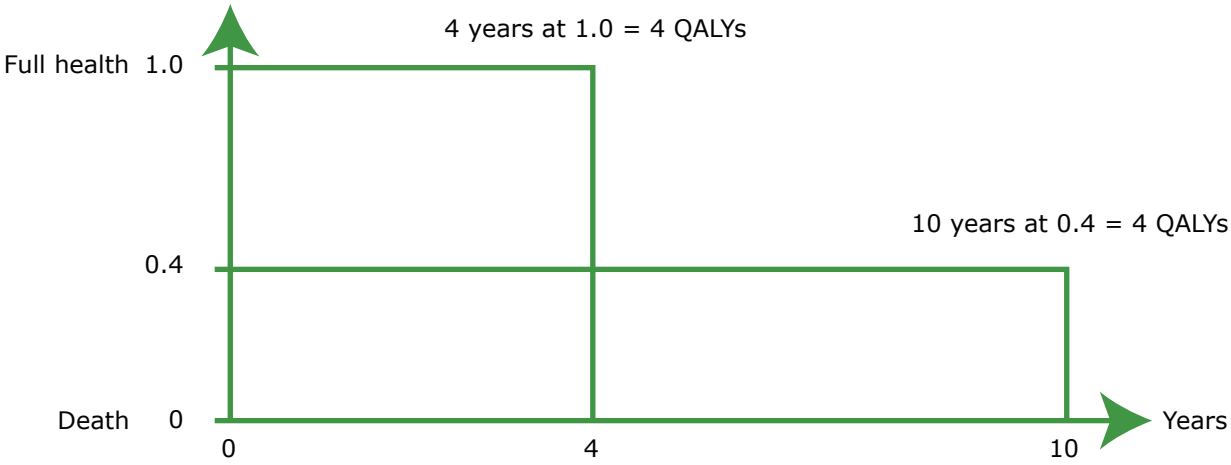
Quality-adjusted survival

Outcome measurement in chronic or progressive diseases—such as Alzheimer's disease, Parkinson's disease, multiple sclerosis (MS), rheumatoid arthritis (RA)—is more difficult, as often no distinct events have an impact on survival. Instead, patients experience a decline in physical and/or mental abilities over time. Often such diseases affect several functions and produce a number of different symptoms, leading researchers to seek an outcome that encompasses all effects. The most frequently used such measure in economic evaluation is the quality-adjusted life year (QALY), which captures the overall effect of a disease on quality of life over a given period of time, and combines the quantity

and quality of life gained from treatment. QALYs can be compared across diseases and thus support choices for resource allocation within an overall health care budget. As a consequence, QALYs are the outcome measure preferred by many government bodies and other authorities that require economic evaluation prior to recommending that a treatment be provided using public funds.

QALYs are calculated by adjusting time (years of life) with an index that expresses global quality of life (utility) on a scale anchored at 0 (death) and 1 (full health). Utility can be measured using techniques from decision analysis that are explained later in this book. For example, if being blind has a utility of 0.4, spending 10 years as a blind person would give four QALYs, which is equivalent to spending four years in full health. Thus, treatments that prolong life (e.g. life-extending cancer treatments) can be assessed in the same way as ones that improve quality of life (e.g. treatments for rheumatic diseases). Figure 2.5 illustrates this concept.

Figure 2.5. The concept of quality-adjusted life years



In order to compare QALYs from different studies, the same methods need to be used in measurement. This is not always done in practice and is one of the reasons why the use of QLAYs has been met with some scepticism.

Without attempting to do justice to the vast literature on QALYs, it is useful to mention the gist of the criticisms made against their use. Some of these centre around the idea that QALYs do not accurately reflect preferences about survival and quality of life. Consider the example described above, where blindness was valued at 0.4. This suggests that the individual achieves four QALYs when this health state is experienced over ten years, eight QALYs when it is experienced over 20 years, and so on. However, it is possible either that some health states become either less tolerable over time—or more tolerable for people who adjust to the condition. For a particular individual, then, 20 years of blindness may seem worth less than, or more than, the QALY value attached to spending ten years in the same state. In addition, some health states may be preferable to death only for a period of time—survival beyond that point may seem less desirable than immediate death. A shorter period of survival in those health states, then, is preferred to a longer period.

From an equity perspective, it is sometimes argued that QALYs discriminate against certain groups, such as the elderly. The potential number of life years that can be saved by treating an 80-year-old patient is fewer than the number of life years that can be saved by treating a 40-year-old patient. This seems to undervalue the elderly patient. Maximising use of QALYs to distribute resources, in addition, implies that all QALYs are of equal social value, no matter who benefits. Society, however, may wish to give priority to certain groups and ensure that those patients have access to treatment even if the cost-per-QALY is high.

Although the QALY is not a perfect measure, its use is widespread because no clearly superior alternative currently exists for making comparisons across diseases. Most decision makers incorporate concerns about QALYs into their decisions by not applying a strict monetary threshold to their willingness to

pay for a QALY. They may be willing to pay more for treatments for certain patient populations or rare diseases. In other words, the cost-per-QALY estimate will not be the only decision criterion.

Monetary outcomes

In cost-benefit analysis, the outcome of a treatment is expressed as the willingness of individuals or society to pay for it. Monetary outcomes have been met with some scepticism in the medical field, mostly due to the reluctance to define a threshold value that society should be paying for a given outcome, such as a life-year or QALY. Furthermore, the techniques for measuring willingness to pay have not been as well tested within the health care environment as techniques for measuring utilities.

Patient-reported outcomes

The interest in measuring patient-reported outcome, i.e. patients' subjective well-being, has increased in recent years. One explanation for this is the increasing number of people with chronic diseases, which predominantly affect patients' quality of life. The objective of treatment here is primarily to improve patients' physical, mental and social functioning. The classical clinical measures are often inadequate for describing and evaluating these effects, so a number of instruments to measure health-related quality of life have been developed, both generic and disease-specific. These instruments are designed to elicit patients' subjective evaluations of the effects of a disease or a treatment and have become an important tool for the assessment of outcomes. However, for the purposes of cost-utility analysis, these measurements can be used only if they are expressed as an index, or weight, with clearly-defined anchors between the "worst" and the "best" health states.

"Health" is defined by the World Health Organisation as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" (WHO, 1948). In general, measurements of health-related quality of life are carried out along these three dimensions. Figure 2.6 lists some of the concepts typically measured.

Figure 2.6. Dimensions in patient-reported outcomes

Dimension	Concept	Includes
Physical	Physical function Symptoms Physical role	Mobility, activities of daily living, self-care Pain, fatigue, nausea Work, household tasks
Mental	Psychological well-being Personal constructs Cognitive functioning	Happiness, depression, anxiety Spirituality, life satisfaction Memory, concentration
Social	Social role Social well-being	Family life, social contacts, friendship Stigma, isolation
Overall	Global judgement of health Satisfaction with care	Overall rating of current health Satisfaction with treatment

Instruments used to measure patient reported outcomes fall into three basic categories, used in different circumstances and for different purposes:

1. Generic measures

2. Disease-specific measures

3. Preference-based measures (utility measures).

Generic measures were developed to assess health status across all diseases and are relevant to all health problems. They have the advantage that the impact of a treatment for one disease can be compared with that of treatment for another disease. A potential drawback of generic instruments is that they may fail to capture small, but important, effects that are specific to a particular disease. To

address this limitation, disease-specific instruments have been developed for many diseases. These measure the distinctive aspects of diseases that are typically missed by generic measures, thus providing valuable information in clinical trials, assessment of specific needs or patient monitoring. However, they are not useful for comparison between diseases and hence cannot be used in decisions relating to resource allocation across therapy areas.

The third category of instrument, preference-based measures, is of particular interest to economists because it yields a set of weights (utilities) on which QALY calculations can be based. Some generic instruments will yield an overall quality-of-life score as an index and therefore can be used as utility measures suitable for generating QALYs. The EuroQol Group's EQ-5D is a typical example of this, as its descriptive "health states" are linked to preference-based assessments. In contrast, another frequently used generic measure, the SF-36, does not produce an overall index, but rather two summary scores for mental and physical domains and therefore cannot directly be used to generate QALYs. More recently, an algorithm extracting domains from the SF-36 to calculate a utility index has been developed (SF-6D) and under certain circumstances can be used to generate QALYs.

All outcome instruments must stand up to scrutiny for reliability, reproducibility, validity, feasibility and sensitivity to change and can be assessed against these criteria using psychometric techniques. Figure 2.7 presents some of the better known instruments.

Figure 2.7. Established outcomes instruments

Type of instrument	Example instruments
General health profiles	Short Form 36 (SF-36), Nottingham Health Profile (NHP), Sickness Impact Profile (SIP), General Well-Being Scale
General health indices	Index of Well-Being, EQ-5D, Health Utilities Index (HUI), SF-6D
Disease-specific scales	Arthritis Impact Measurement Scale (AIMS), Minnesota Living with Heart Disease Scale, Multiple Sclerosis Quality of Life Inventory (MSQLI), Beck Depression Inventory (BDI)

The development of a quality-of-life instrument is a complicated process that can span several years. The Medical Outcomes Study SF-36, for example, was developed over a period of ten years, using questionnaires and data from the RAND Medical Outcomes Study in the US. It was translated, adapted and validated in a large number of countries. Acceptability of a new instrument will depend on its use in several different investigations, adding further delay to its widespread use. Thus, development of new disease-specific instruments should be undertaken only when no adequate instrument is available and this lack of availability cannot be overcome by, for instance, using a combination of existing instruments that together address the concepts required.

Cost Data for Economic Evaluation

Perspectives

In order to capture all costs that are of relevance to society, economic evaluations should be performed from a societal perspective. However, a number of jurisdictions make decisions about new treatments from the perspective a public payer, covering health care, social services, and pensions, or a health care payer, covering only health care and related services. The perspective used will determine which resources are included in the analysis.

A societal perspective includes all costs, regardless of who incurs them. Thus, costs to the health care service, social services, patients and the rest of society (for example, in the form of production losses) are included, but transfer payments are ignored. Examples of transfer payments are taxes and reimbursement for income loss due to illness. For society as a whole, taxes and reimbursement represent a money flow from one part of society to another, but no resources (labour, capital) are being used up. The relevant concept of cost in economics is that of opportunity cost, i.e. the benefit foregone

from using resources for one purpose, rather than for their best alternative use. This definition serves to remind us that costs will be incurred even when the use of a resource is not associated with any financial flows, such as in the case of a voluntary caregiver.

From the perspective of a third-party payer—e.g. a government, insurance company, or managed care organisation—only resources paid for by that organisation are included as costs. For instance, any reimbursement to patients for income loss is an actual cost to the third-party payer. A good example of the effect of different perspectives is shown in Table 2.1., taken from a cost of illness study for MS. In the perspective of the health care payer, only medical costs are included.

Table 2.1. Mean costs per MS patient in Germany

Cost per person and year (2005 EUR)		
Type of Cost	**Societal perspective**	**Public payer perspective**
Inpatient care	3,203	3,133
Consultations	3,096	1,860
Tests	368	368
Pharmaceuticals	10,498	9,588
Services	525	241
Adoptions (investments)	989	393
Informal care	4,407	—
Production losses	16,911	—
Transfer costs (pensions)	—	3,404
Total costs	*39,998*	*18,988*

Source: Kobelt et al (2006a)

Steps in cost assessment

Assessing the costs in an economic evaluation involves four steps, which are identical in all forms of economic analyses.

1. Identify the relevant resources used
2. Quantify these resources in physical units, such as hospital days, admissions, surgical procedures, physician visit, tests, etc.
3. Value the different resources used in terms of their opportunity costs
4. Adjust valuations to account for the differential timing at which resource use can occur (discounting)

Identification of resources. Relevant resources will be defined by the study objective. In a cost-of-illness study, this will include all resources related to the disease, its consequences and its treatment. In an economic evaluation, relevant resources can be defined as those that are related to the administration and consequences of the treatment during the disease episode concerned. For example, if two different surgical interventions for the same problem are to be compared, such as open surgery and laparoscopic surgery, resources related to the original disease diagnosis are not relevant, as these are identical for both alternatives.

Resource quantification. The way in which resources are quantified will depend on what needs to be measured and whether a unit cost can be assigned to it. If an intervention reduces hospital days, one will logically collect hospitalisation data in the form of length of stay. If it reduces the number of hospital

admissions, one would collect data on admissions (or discharges). But if the intensity of care within the hospitalisation is reduced by the intervention, it will be necessary to collect all details on resources used during the stay.

Figure 2.8. Typical items of resource use in an economic evaluation

Cost type	Examples of resources
Direct medical costs	Hospitalisation • Days of hospitalisation • Discharges Outpatient visits • Outpatient clinic attendance • Visit to private practitioner • Visit to paramedic Procedures and tests • Tests (blood analysis, X-ray, scans, gastroscopies, etc.) • Surgical interventions Devices • Medical devices (wheelchairs, hearing aid, pacemakers, etc.) Services • Home care (hours or days) • Nursing care (hours or days)
Direct non-medical costs	Transportation • For outpatient visits (ambulance, taxi, etc.) • For daily activities Services • Home help (hours or days) • Meals on wheels • Social assistance (hours or days) Devices and investments • Adaptation to house or car • Special kitchen and bathroom utensils Informal care • Care by relatives (is sometimes also considered an indirect cost)
Indirect costs	Sick leave (days or weeks) Reduced productivity while at work (percentage or hours) Early retirement due to illness (years to normal retirement) Premature death (years to normal retirement)

Resource valuation. The quantity of units used is multiplied by their unit cost (price) to obtain the total cost. The way in which resources have been quantified will determine what unit costs are assigned to them. Admissions or discharges will be costed using aggregate measures of resource use (macro-costing), while costs incurred during the hospitalisation itself will require unit costs for each individual resource (micro-costing).

Figure 2.9. Dimensions of costs and prices

Opportunity costs	Cost of the next best alternative foregone
Tariffs	Price defined by or negotiated with a third party payer
Charges	Billings to third party payers or patients

Unit costs should represent opportunity costs (Figure 2.9). Bearing in mind the concept of opportunity cost as benefit foregone, a simple example of opportunity cost is that of a physician's time during a consultation. The time used during one consultation cannot be used for another consultation, and hence has a cost. The opportunity cost in this case is the value lost for the consultation that was not undertaken. In normal well-functioning markets, market prices provide a good representation of the opportunity costs of resources, but in health care this is not always the case. In countries with a national health service, such as the UK or Sweden, resources may not be subject to market valuations.

In some countries, the only easy source of costs is tariffs, i.e. prices set by a government or a public insurer for payment to health care providers such as hospitals or physicians. In many cases, however, tariffs do not represent the actual opportunity costs. In fee-for-service systems where each service is paid for separately, tariffs may be set to include incentives for the level of supply of a given resource, with high tariffs set to encourage provision and low tariffs to discourage it. An example is shown in Table 2.2, which summarises a study of the cost of glaucoma in Germany.

Table 2.2. Tariffs and opportunity costs in a cost of illness study (glaucoma) in Germany

	DM per unit (1997)	
Resources	**Insurance tariff (quarterly billing)**	**Opportunity cost (time, supplies, overheads)**
Consultations		
• **first visit/quarter**	19.11	34.62
• **subsequent visits/quarter**	3.56	34.62
• **telephone**	3.56	3.68
Tests		
• **Goldman**	0.00	13.71
• **periscopy**	28.48	12.64
• **gonioscopy**	9.26	4.68
• **ophthalmoscopy**	0.00	5.20
Outpatient procedures		
• **trabeculectomy**	149.52	377.85
• **laser trabeculectomy**	71.20	89.65

Source: Kobelt et al (1998)

In other countries, the most readily available unit costs may be billings (charges) from providers to different payers, generally insurers or health plans. Such charges often are used to subsidise other activities, e.g. within the hospital, and will hence be higher than the opportunity costs. This is the case for instance in the US, where a cost-to-charge ratio of 1:2 is often applied.

Challenges arise in applying appropriate valuations to resources that have an opportunity cost, but no clear market price, such as informal care by family members or friends. These costs can be important in disabling diseases and chronic diseases prevalent in the elderly, in particular. Agreement has not yet developed on whether and how to include such resources in economic evaluation. Two methods are generally used: "replacement cost" (in this case, the cost of a professional providing the care in lieu of the family), or the loss of leisure time while providing care, commonly valued as disposable income. However, these costs do not necessarily have to be valued in monetary terms for decision makers to take them into account.

The role played by indirect costs (production losses) will to some extent depend on the pathology being analysed. In diseases such as asthma, depression, schizophrenia, MS and migraine, indirect costs tend to make up a sizeable proportion of the total cost of the illness because these diseases affect age groups with high labour force participation. In diseases that affect predominantly elderly people, indirect costs would be less important.

Approaches to the valuation of indirect costs differ. In general, valuation is based on human capital theory whereby an individual's value is based on their "market price", in this case the total cost of employment, which is gross salary plus employers' contribution. Time lost from work—sick leave, early retirement or premature death—is estimated based on the average national salary adjusted for age and sex, if relevant.

Some believe that the human capital method overestimates indirect costs. An alternative, the "friction-cost method" has been proposed. This asserts that production losses are limited in times of high unemployment as workers who retire early are replaced more readily, although this assumes that people with the right skills and qualifications are available. For short-term absences, the friction-cost method sees lower losses in production due to sick leave because work can be temporarily redistributed to other employees, although this assumes that spare capacity is available. This method also seems to ignore that the short-term cost of compensating for lost output due to absence from work includes, for example, maintaining spare capacity or paying more for overtime work. Thus, under to this approach, lost production capacity due to temporary absences or early retirement is largely ignored, producing considerable differences in estimates based on the two methods.

Where the results of the two methods diverge substantially is in the estimate of production losses from mortality and long term disability. Under the human capital approach, production losses are estimated over the entire period of lost employment. In contrast, the friction-cost method estimates production losses only for a limited period (the friction period), after which absent employees are replaced. Table 2.7 illustrates the difference between the two approaches using a study of RA in The Netherlands.

Table 2.3. Difference in production loss estimates using the human capital and the friction-cost methods

Mean annual indirect costs (2005 EUR) in a sample of patients <65 years old with RA			
	Men (n=91)	Women (n=261)	Sample (n=352)
Friction-cost method	827	325	455
Human capital method	12,789	5,125	7,109
• **Sick leave**	-1,172	-482	-660
• **Reduced working hours**	-1,648	-157	-542
• **Early retirement**	-9,978	-4,487	-5,905
Difference	11,972	4,800	6,654

Source: Verstappen et al (2005)

As mentioned earlier, authorities differ by jurisdiction in their willingness to accept indirect costs as part of economic evaluation. Descriptive studies such as cost-of-illness studies always should include indirect costs. All basic evaluations should be from a societal perspective and so include indirect costs. It is easier to exclude such calculations from discussions where they are not needed, or accepted, than it is to estimate them later on, after the evaluation has been completed.

Cost-of-Illness Studies

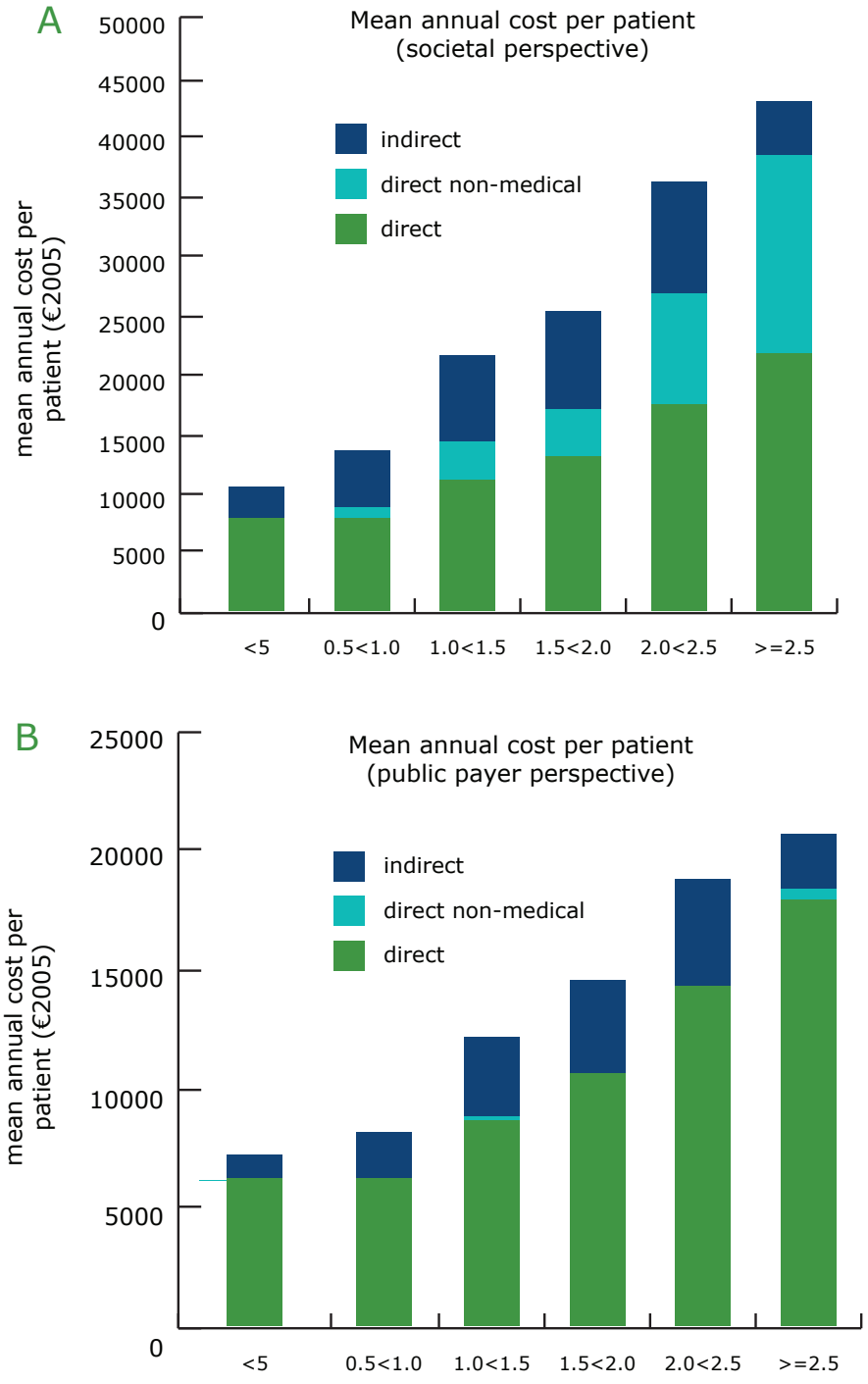
Cost-of-illness (or burden-of-illness) studies are not concerned with a particular health care intervention, but instead attempt to estimate the economic burden that a specific disease places on society. They are not good guides for resource allocation. No matter how great the cost of a disease, devoting resources to it serves no purpose if no effective treatment is available. Moreover, because such studies do not assess improvement in health from a specific intervention, they cannot indicate where resources should be invested to achieve the most health gain.

Cost-of-illness analyses, then, are not economic evaluations, but instead act as points of reference for economic analysis. Most are limited to estimating direct and indirect costs, although intangible costs are sometimes calculated. Costs can be analysed based on prevalence or incidence.

Prevalence-based studies

In prevalence-based studies, all costs are estimated for a patient population in a given geographical area for a given period of time, generally one year. Such studies are useful to health policy makers for planning and budget decisions. For example, a study might estimate the amount that a given country spends per year on caring for patients with Alzheimer's disease. As more people live longer, the number of patients with Alzheimer's disease will increase; a prevalence study of Alzheimer's disease, then, can help plan for future demand and cost. If analyses for several years are available, changes in cost over time can be calculated and may help forecast future trends. An example is shown in Table 2.4, which illustrates how the costs of cancer in Sweden have changed over time.

Figure 2.10. Association of severity of rheumatoid arthritis and severity of disease



Note: Severity of RA was measured using the Health Assessment Questionnaire
Source: Kobelt et al (2008b)

Table 2.4. Prevalence estimates of the cost of cancer in Sweden (2009 EUR mn)

Year	Direct costs as % total	Indirect costs as % total		Total costs
		Morbidity	Mortality	
1975	22%	20%	58%	544
1983	29%	17%	54%	1,120
2004	50%	11%	39%	3,600

Source: Jönsson and Karlsson (1990); Wilking et al. (2010)

Cost-of-illness studies also can demonstrate how costs are distributed across resources and where major expenses occur. Figure 2.10 illustrates this using the cost of RA in France.

Cost-of-illness analyses can help national policy makers gain insight into where the country's health care resources are being spent. Costs also can be compared across countries if studies using the same methodology are performed in countries with similar economic conditions, as shown in Study Example 1. Such comparisons, however, are not always straightforward. They may need to be adjusted to take account of the effect on costs of differences in management strategies, resource utilisation, unit costs, payment mechanisms, and even the characteristics of the sample of patients included in the studies. Results in one country, therefore, seldom are applicable in other countries.

Study Example 1. Prevalence-based cost of illness—multiple sclerosis

Multiple sclerosis is an autoimmune disease that affects young adults and rapidly leads to severe physical disability. Over the past 15 years, the introduction of several expensive new treatments aimed at slowing progression of MS has focused attention on current and potential future expenditure. A considerable number of studies of cost have been performed. Studies in the early 1990s generally found that indirect costs constituted the vast majority of costs (70–80%). Many of these studies were small and did not capture some costs that fell outside the health care system (e.g. costs incurred by patients, costs of informal care). Most importantly, only limited information on how costs and quality of life evolve with advancing disease was available. Such information is important in a setting where treatments aim to delay progression to severe disability.

A series of observational studies in three countries collected information about resource use, quality of life and disease parameters directly from population-based samples of patients, allowing extrapolation to total costs in each country. These are shown in Table 2.5.

Table 2.5. Cost of MS in three countries (2000 EUR)

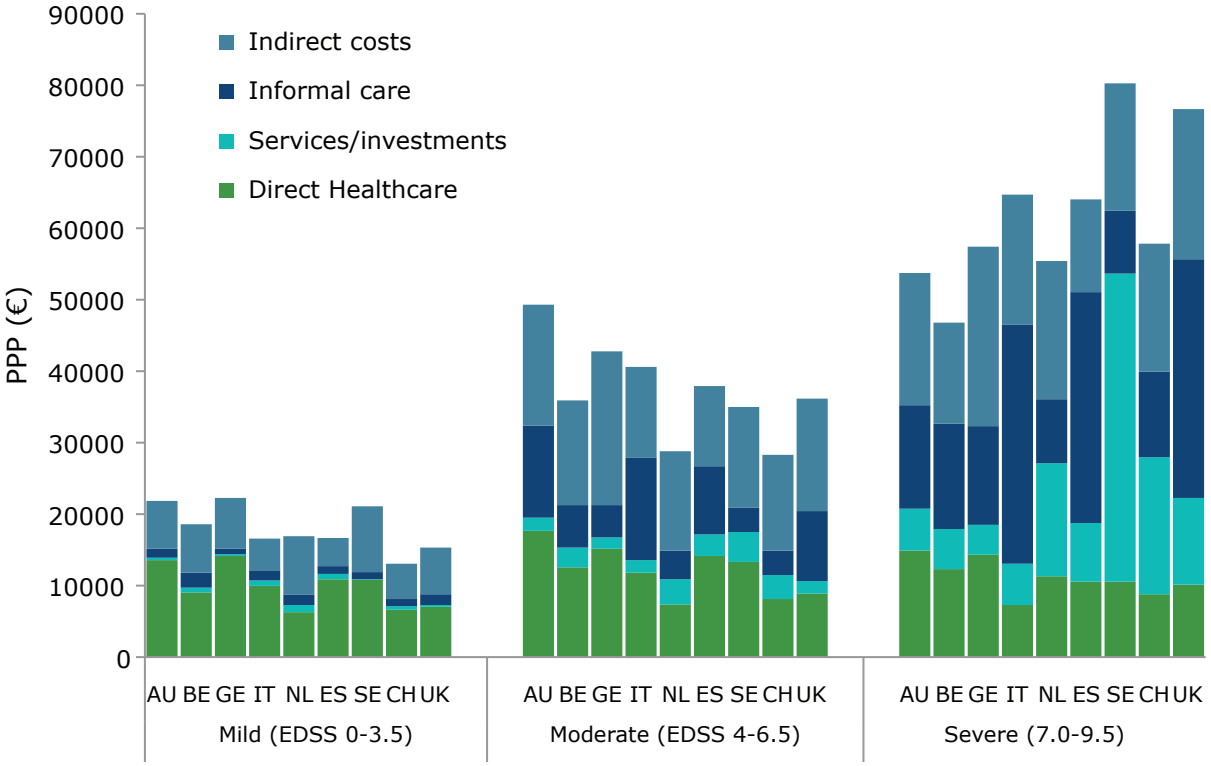
	Sweden	UK	Germany
Estimated prevalence	11,000	88,000	120,000
Cost per MS case	45,000	28,000	33,500
Total estimated cost of MS	0.5 bn	2.2 bn	4.0 bn
Cost per inhabitant	56	36	50

Source: Henriksson et al (2001); Kobelt et al (2000); Kobelt et al (2001)

More recently, a large series of observational studies including over 15,000 patients with MS was performed in ten countries across Europe (Kobelt et al, 2006b). The studies collected information on all medical and non-medical resource consumption, services, devices and investments, informal care by relatives, and production losses along with data on relapses, disease severity and overall quality of life (elicited directly from patients using questionnaires). Although the study method was identical in all countries, direct comparisons of the overall results are not possible. This is because the samples may present different severity profiles, which will drive differences in consumption. However, the dataset allows resource consumption (costs) to be correlated with disease severity (measured in MS

generally with the EDSS, a disability status scale with scores ranging from 0 to 10) and with utility, as shown in Figures 2.11 and 2.12.

Figure 2.11. Association of cost and severity of multiple sclerosis



Source: Kobelt et al (2006b)

In this case, the data permit some cost comparisons across countries at given levels of disease severity, although differences in unit costs must be taken into account. It is also possible to draw some conclusions regarding the more general development of costs, e.g. with advancing disease. For example, production losses increase dramatically from mild to moderate disease, indicating that patients with moderate disease have to leave the workforce. The largest cost increases are found in services and informal care, as disease severity worsens. It is also in these areas that the largest differences in strategies across countries can be observed: the more services a country provides, the greater the reduction in the need for informal care, as is demonstrated by Sweden, Switzerland and The Netherlands.

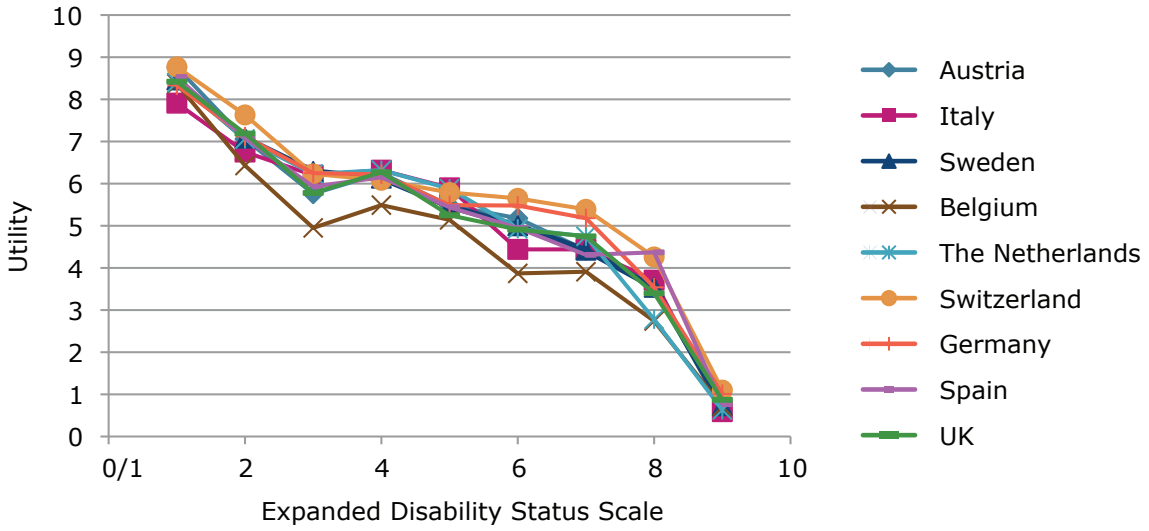
In Figure 2.12, the overlapping lines suggest that utilities are more "universal" than costs. While resources need to be collected for every country and valued using that country's unit costs, utilities from one country may sometimes be used in an economic evaluation conducted in another country. However, note that some agencies, such as NICE, demand that only data from their own jurisdictions be used.

As the studies in this example included all types of resources (patient and carer costs, in particular, which were generally omitted in earlier studies), total costs were greater than had previously been found, while the proportion of costs represented by production losses was much smaller.

Several issues need to be considered when collecting information directly from patients. Patients' ability to recall events is not perfect and they may overstate or understate resource use. Also, patients may find it difficult to distinguish between costs that are related directly to the disease and those that are not. The latter is a particular problem in diseases areas where general medical practitioners

(primary care physicians) are involved in providing care. However, in MS, this was not considered a major problem, as patients are generally young, so co-morbidity is limited.

Figure 2.12. Relationship between utility and severity of multiple sclerosis



Source: Kobelt et al (2006b)

Accuracy of recall is a difficulty in all areas and should be taken into account when planning the timing of data collection. While a recall period of twelve months for hospital admissions or major investments has been shown to be reasonable, the lag time in collection of information regarding consultations, tests and sick leave should not exceed three months. For smaller items such as drug usage (other than treatments taken regularly over a long period of time), even three months is considered too long, and one month is more appropriate. When time is concerned, such as the number of hours of care given by family members or number of hours of reduced productivity, it is advisable to use even shorter recall periods. Indeed, a widely used instrument to measure productivity while at work or during other activities (Work Productivity and Activity Index) covers just one week. When modulating recall periods in this way, some control mechanism or testing should be included in the study. In the series of three observational studies in MS described earlier, hospital records were compared to patients' answers, and it appeared that there was no recall bias. In Germany, for instance, the mean number of inpatient days according to hospital records were virtually the same as those reported by patients (a mean of 26.9 and 27.15 days, respectively, with similar ranges).

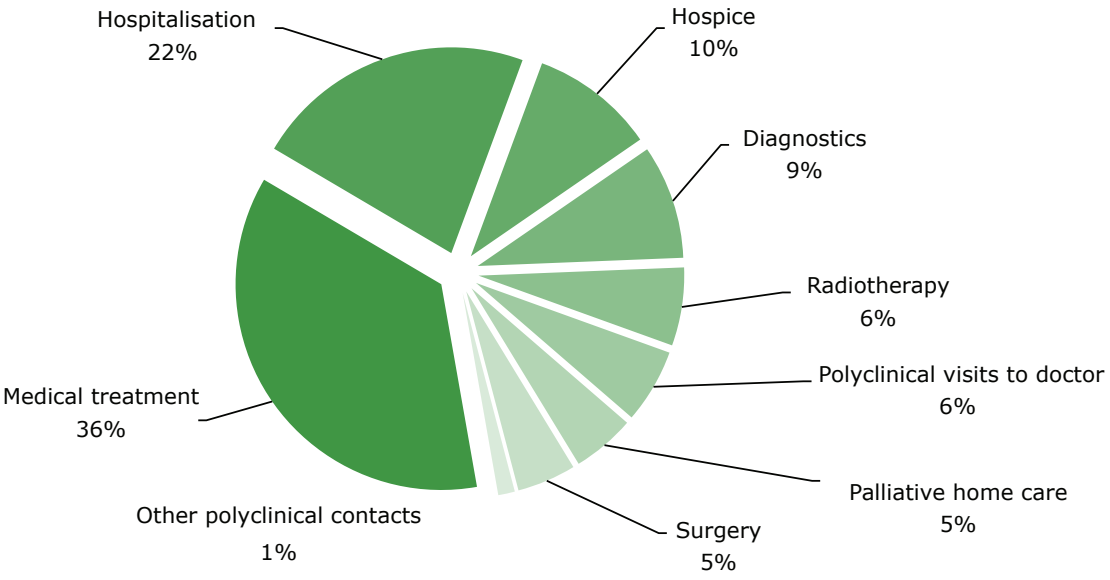
Incidence-based studies

In incidence-based studies, lifetime costs for a patient with the disease are estimated—from diagnosis to cure or, in chronic diseases, from diagnosis until death. These studies are more useful than cost-of-illness studies when estimating the effect of a treatment on future costs. For Alzheimer's disease, for example, incidence studies may identify costs such as nursing care that could be avoided by treatment that prevents the loss of mental capacity.

Incidence-based studies of chronic diseases that span decades are difficult to perform in comparison to terminal illnesses for which it is possible to work backward from death to estimate costs. Studies of chronic disease often are limited to examining costs per case over a given number of years to identify what treatment strategies prevail and what drives the costs (see Study Example 2 below).

Figure 2.13 presents an example of an incidence study of direct medical costs from diagnosis to death for metastatic breast cancer.

Figure 2.13. Direct medical costs for metastatic breast cancer



Source: Dahlberg, Lundkvist and Lindman (2009)

Costing approaches

Data for cost-of-illness studies can be identified from a range of sources such as national health care statistics, patient registries, cohort studies, insurance databases, patient charts (as in Study Example 2) or from patients themselves (as in Study Example 1). Depending on the availability of data and the level of detail required to answer the study question, studies are performed either "top down" or "bottom up".

- In *top-down studies,* statistical databases and registries are used to estimate the costs for a given prevalence sample, providing aggregate data at the regional or national level. The problem with this approach is that, in most countries, some costs are not available from these source and total costs thus will be underestimated. Moreover, databases may be incomplete or items miscoded.

- In *bottom-up studies,* costs are collected directly from a patient sample, either retrospectively using patient charts and questionnaires, or prospectively by following the sample for a period of time. The results for the sample then are extrapolated using prevalence data to estimate costs at a regional or national level. The difficulty with this approach is ensuring that the sample is unbiased and representative of the overall patient population. Recall bias also may raise issues, as noted earlier.

Table 2.6 illustrates the differences between top-down and bottom-up cost-of-illness studies in Germany from the perspective of the payer and society.

Table 2.6. Differences in national costs between top-down and bottom-up cost-of-illness studies in Germany from different perspectives (prevalence 120,000)

	Top Down*	Bottom up**	
	Public Payer Perspective	Public Payer Perspective	Societal Perspective
Direct costs	1,031	2,047	4,525
Inpatient care	353	750	804
Ambulatory care	437	477	556
Drugs (interferons)	143	344	365
Drugs (other)	75	154	186
Services, adaptations	23	322	1,664
Informal Care	—	—	950
Indirect costs	412	930	3,349
Sickness absence	133	50	296
Early retirement	288	880	3044
Total costs (DM)	*1,432*	*2,977*	*7,850*
Total costs (EUR)	*730*	*1,520*	*4,010*

*1998 **1999

Source: Upmeier and Miltenburger (2000); Kobelt et al (2001)

Study Example 2. Incidence-based cost of illness—glaucoma

Glaucoma mainly affects the elderly and is characterised by a gradual restriction of the visual field due to damage to the optic nerve, potentially leading to blindness. The causes and progression of the disease are not understood fully, but elevated intraocular pressure (IOP) is considered to be the major risk factor and hence is the main target of all treatments, both pharmacological and surgical.

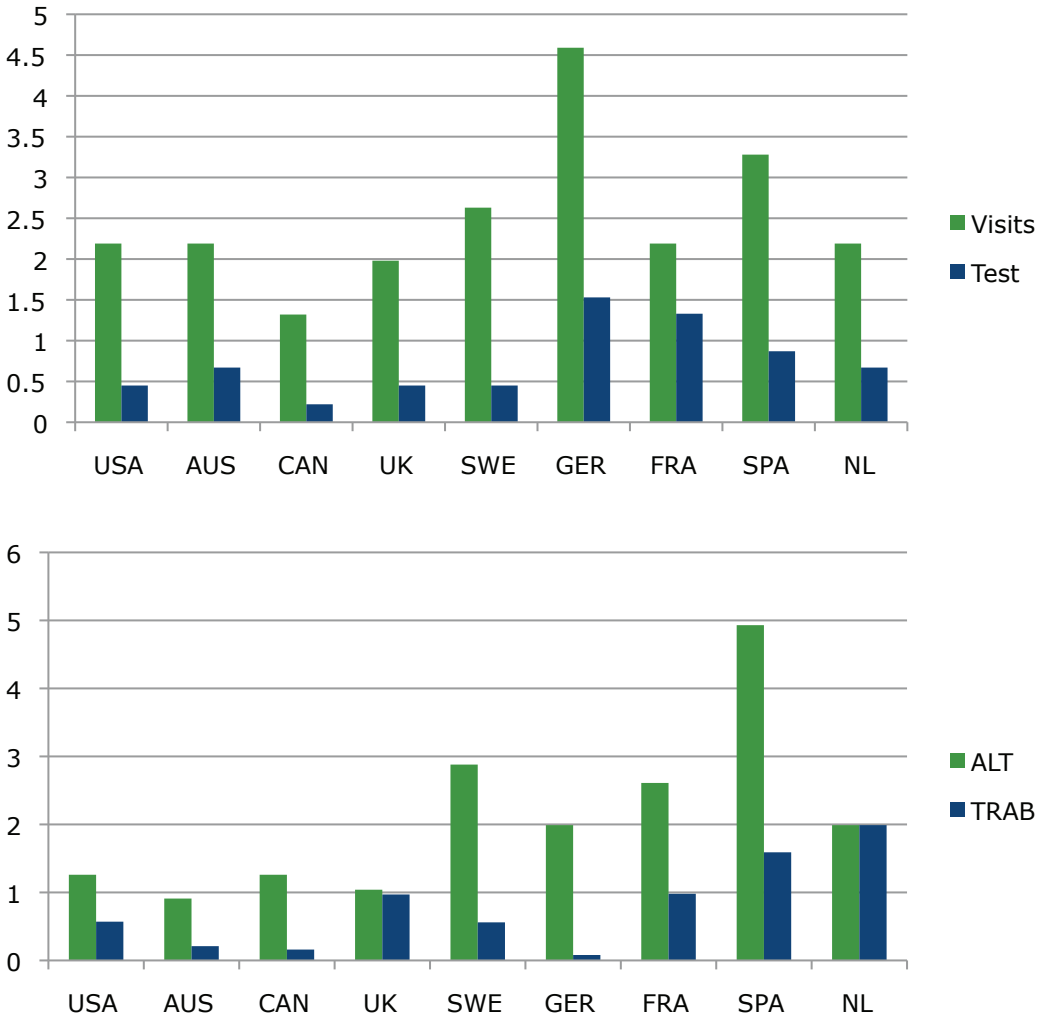
A prevalence-based study to investigate annual spending on the disease would be useful to forecast the increase in expenditure as the population ages. However, if the study is intended to provide the basis for a cost-effectiveness analysis of a new treatment, it is important to investigate how treatment patterns and costs develop over time for individual patients or patient groups. Thus, a longitudinal study of newly diagnosed patients will be more useful, even if it does not cover the entire time span from diagnosis to death.

A study by Jönsson and Krieglstein (1998) was performed as a retrospective chart review in nine countries covering the first two years of treatment after diagnosis. The main purpose was to establish a baseline of current clinical practice to estimate the impact of the introduction of a new therapy. Specifically, the analysis included patients on standard treatment who would qualify, according to clinical judgment, for the new therapy. The study investigated the time to failure of first-line therapy and the preferred treatment strategies thereafter, as well as the major drivers of costs and differences between countries.

Study sites were selected based on the organisation of ophthalmic care in each country. Patient files were searched from December 1995 backwards, including all patients with complete two-year data, until a sample of at least 200 was reached. Medical data were limited to detailed diagnosis and development of IOP over two years, but all resource utilisation data related to glaucoma were included. Resources were valued at their opportunity costs, but the data were limited to direct health care consumption.

The study found that medical parameters were surprisingly similar. For instance, in all nine countries, the mean post-treatment IOP was 18 mmHg, despite the fact that the mean at diagnosis varied from 31 mmHg (Germany and Sweden) to 24 mmHg (France). However, the treatment paths to reach this target level were vastly different, as can be seen in Figure 2.14. As a consequence, large differences in costs are apparent across countries, although it again should be borne in mind that such a comparison has to be interpreted with care due to the differences in prices and health care organisation (as described earlier, and as illustrated in Figure 2.14).

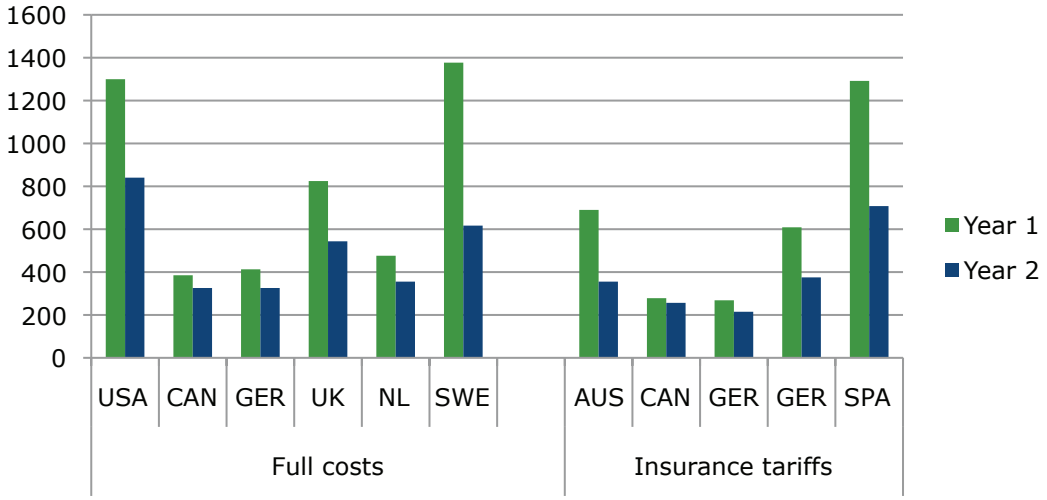
Figure 2.14. Differences in the choice of treatment and patient management across countries



Source: Jönsson and Krieglstein (1998)

Multiple regression analysis identified the main cost drivers as IOP and change in IOP with treatment. In all countries, overall costs increased by approximately 4% for each mmHg unit increase in IOP at diagnosis; costs decreased by approximately 3% for each mmHg unit decrease in IOP with treatment. This was due to more intensive management and more frequent treatment changes when IOP was higher. Thus, in all countries, costs increased with each treatment change, illustrated in Table 2.7 for the UK and US. The treatment sequences and cost drivers identified in this study served as basis for a model to estimate the impact of a new drug (see Study Example 4).

Figure 2.15. Differences in costs depending on the use of estimates based on insurance tariffs or full opportunity costs



Source: Jönsson and Krieglstein (1998)

Table 2.7. Mean two-year costs per patient by the number of therapy changes

	UK		USA	
Number of Changes	**% of patients**	**Costs (GBP)**	**% of patients**	**Costs (USD)**
0	73	372	53	1,424
1	15	1,346	18	2,121
2	8	2,788	13	2,950
3 or more	4	2,834	16	4,458

Source: Jönsson and Krieglstein (1998)

This study highlights some of the advantages and disadvantages of retrospective data collection. Analysing clinical records retrospectively has obvious limitations, such as the lack of control as to how the data were obtained and, more importantly, missing data. Thus, prospective data collection may sometimes be preferable to ensure that all information is available and in an appropriate format. However, for resource utilisation, a retrospective design has certain advantages: the data have not been influenced by any protocol or study design and thus represent true clinical practice, and the study can be carried out in a relatively short time with limited resources.

Chapter 3

ANALYTICAL APPROACHES TO ECONOMIC EVALUATION

In principle, the most straightforward way to estimate costs and consequences is to use resource utilisation and efficacy data from randomised clinical trials. This approach retains the high internal validity of the trial, ensures that both costs and effects are measured in the same setting, and allows variability in the estimates to be explored using confidence intervals for the ICER.

However, this approach often is not suitable in practice for several reasons. First, in many diseases it is impossible to enrol enough subjects for a long enough time period to collect the necessary resource data. Second, the special circumstances of clinical studies will influence patient management and some costs will be entirely protocol-driven, preventing a relevant comparison to clinical practice. For example, if a clinical trial protocol includes a systematic test that in clinical practice would only be performed if needed by clinically apparent symptoms, investigators will have to act on the test results. This will produce higher resource consumption than in clinical practice. Third, many studies enrol patients in a large number of countries and the individual national groups are often too small to assess country-specific costs, as would be needed for an economic evaluation. One way of handling this issue has been to use the quantities of resources from the entire trial and apply country-specific unit costs to them. However, because this does not take into account the fact that patient management may differ across countries, further adjustments may be required.

The difficulties in using trial data for economic evaluation are not limited to resource utilisation. Participation in a clinical trial, particularly in registration trials, tends to be restricted to a narrowly defined group of patients in order to isolate the effect of the treatment in a particular set of circumstances. Thus, trial results may not reflect what is likely to happen to the broader population of patients with the condition being treated. Furthermore, trials that compare a treatment with placebo will not deliver any of the comparative data required by the decision maker to evaluate what improvements in health and changes in resource consumption will occur when using a new treatment in place of existing standard therapy.

These characteristics of clinical trials usually mean that efficacy results are of limited generalizability beyond the trial and that it is difficult to translate such results into estimates of effectiveness in routine clinical practice. A number of statistical approaches to address some of these issues, and the uncertainty in the data, have been developed. Some are discussed in this chapter. It should not be forgotten, however, that no statistical method can make up for an absence of adequate data.

Under the circumstances, the generally accepted approach is to model costs and effects by synthesising data from different sources (epidemiological, clinical, economic). Models are created to structure the decision problem in a logical framework to support the decision. Within models, it is possible to combine different data sets, extrapolate to a longer time frame than clinical trials, and test different assumptions about elements such as risk, effectiveness and costs.

Economic evaluations generally use one of the following types of model structure.

1. Decision tree models are used for cost-effectiveness analysis in diseases with distinct events that occur with a given probability, either by decision or by chance, within a relatively limited time frame

2. Markov models or discrete event simulations are more appropriate for analyses in diseases with an ongoing risk over a long time period.

Decision Analysis and Modelling Techniques

Decision analysis was developed as a discipline for examining choices under uncertainty and has long been applied to clinical decision making. It enables complex problems and processes to be broken down into component parts, each of which can be analysed individually in detail before they are recombined in a logical, quantitative and temporal way to indicate the best course of action.

Decision trees

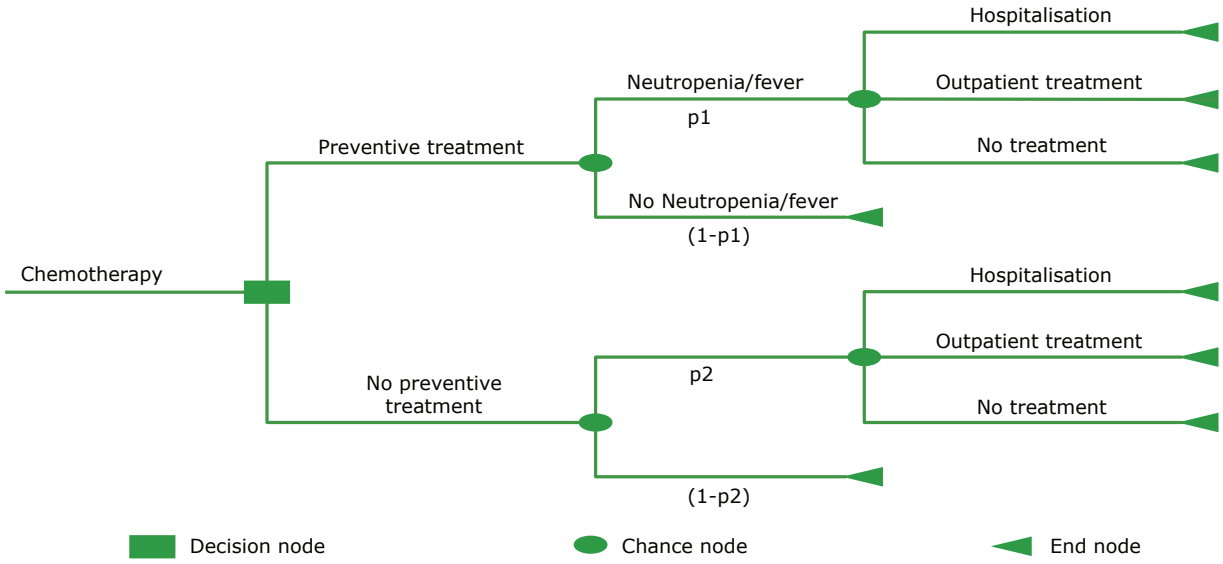
Analyses can be depicted as decision trees that incorporate strategic choices, probabilities of subsequent events and final outcomes. An example is shown in Figure 3.1.

Several steps are required to construct a decision tree: clear definition of the problem, description of successful or unsuccessful outcomes, definition of alternative patient management strategies and their consequences, estimation of the probabilities, and a time frame. Decision trees are usually based on data from clinical trials and other sources of empirical evidence, such as systematic reviews and meta-analyses. For the economic evaluation, the expected cost for each strategy is calculated by multiplying the cost for each branch by the overall probability of that branch occurring. The expected outcome is calculated in a similar fashion by multiplying the defined outcome (e.g. cure, event) by the overall probability of that outcome occurring. The different treatment strategies then can be compared in terms of their different expected costs and outcomes.

Box 3.1. Decision tree

In the example of a decision tree in Figure 3.1, a decision (decision node) is made to give or not to give a treatment that reduces the risk of chemotherapy-induced neutropenia. In both cases, patients may suffer neutropenia, but the probability (chance node) in the treatment group (p1) is lower than in the no treatment group (p2). Consequently, the cost of treating neutropenia is lower in the treatment group, as fewer patients have neutropenia, assuming that both groups are treated for neutropenia in the same way. Expected costs and expected outcomes for treatment or no treatment are estimated by summing up all the branches ("folding back the tree").

Figure 3.1. Example of a decision tree



Using the decision tree model above, if we assume that the preventive treatment costs €1,000, that the average proportion of patients experiencing neutropenia without prevention is 40%, that treatment reduces this risk by 25%, and that the average cost of treating a neutropenic event is €3,000, then the average cost per patient would be €1,900 (€1,000 + (3,000×0.3)) in the prevention arm, and €1,200 (€3,000×0.4) in the no-prevention arm.

The cost-effectiveness of preventive treatment can be estimated by comparing the two strategies. In this example, the incremental cost per neutropenic event avoided would be €7,000 (€700/0.1). In other words, preventive treatment would reduce the absolute proportion of patients with neutropenic events by 10%, thereby saving €300 (€3,000×0.1) and leaving an incremental cost for the preventive treatment of €700 (€1,000–€300).

This example is limited for illustrative purposes and it has to be borne in mind that an outcome such as "neutropenic event avoided" is intermediary and hence of limited use to decision makers. Funding decisions will require information on the consequences of avoiding such an event (e.g. impact on survival or QALYs).

Markov chain analysis

Decision trees often will not be the best way to describe disease effects and interventions. This is particularly the case in chronic diseases where the risk of disease progression, for example, may be continuously changing over time and where events and their timing are important features of the disease. In such cases, a Markov model or discrete event simulation will be more appropriate. Box 3.2 describes the structure of Markov models and Figure 3.2 illustrates the structure of Markov models.

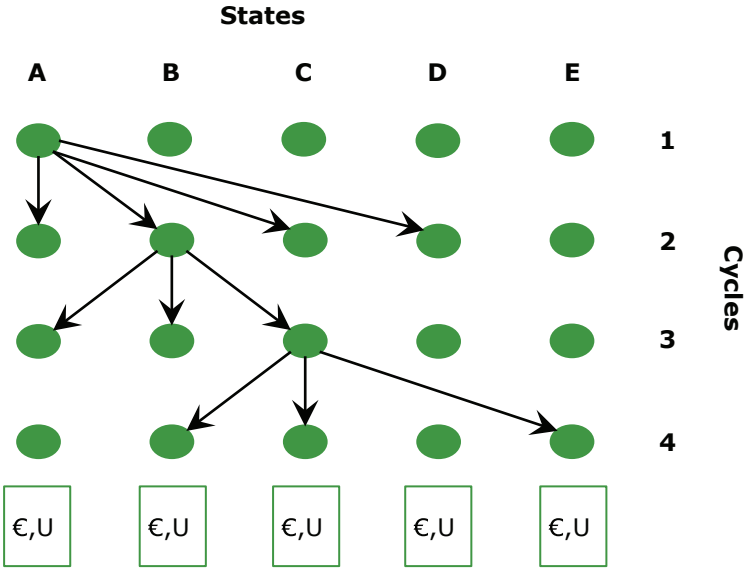
For Markov models, it is assumed that all patients (cohort) can be classified into a finite number of mutually exclusive states, so called "Markov states". These states generally are defined by disease parameters, such as level of severity, or by health states with defined symptoms that are meaningful to patients and clinicians, but other definitions exist as well, such as being on a given treatment. Development of a disease and the effect of treatment are represented as transitions from one state to another. Disease progression will be represented by transitions to more severe states, while the treatment effect will either reverse or slow this progression. The differences or cut-off points between the states must therefore also represent clinically meaningful differences.

Box 3.2. Structure of a Markov model

Markov models illustrate the disease process by distributing all patients across a finite number of distinct and mutually exclusive disease states at baseline and then following the development of the cohort during a defined time (number of cycles).

Figure 3.2. Illustration of a Markov chain analysis

- All patients are grouped into a finite number of states (Markov states)
- Time progresses in equal increments (Markov cycles)
- All events or progression are represented as transitions from one state to the other, with a certain probability
- Spending one cycle in a given state is associated with a definited cost (EUR) and a defined utility (U)



For instance, states could be defined by levels of disability, with state A above being "no disability", B "mild disability", C "moderate disability", D "severe disability" and E "death" (absorbing state). All Markov models require a state that patients cannot leave, usually death, in order to perform survival analyses. However, often not enough detailed information is available to perform lifetime analyses. In those cases, the duration of the model—i.e. the number of cycles combined with their length—is chosen based on the disease, the epidemiological and clinical data that are available, and the economic question to be answered.

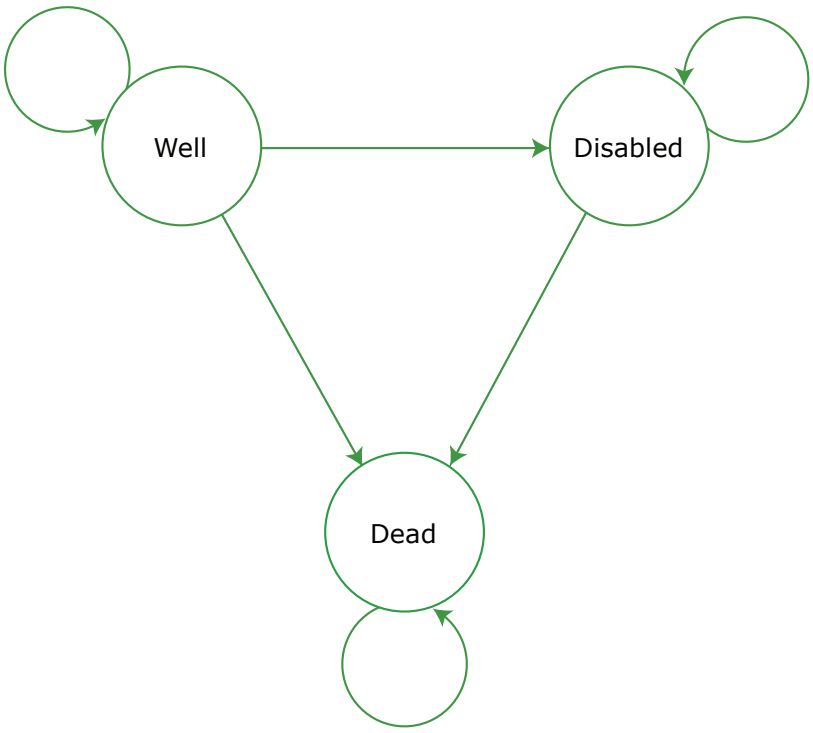
Costs and utilities (health status) for these states are assumed to depend on the state only and are therefore the same for all cycles. In such a framework, more severe states are generally associated with higher disease costs and a lower quality of life. Thus, if patients spend more time in the benign states of "no disability" or "mild disability", costs within a given time frame will be reduced while quality of life will be improved.

The transitions between states, i.e. the probability at each cycle of a deterioration (e.g. from moderate to severe disability) or an improvement (e.g. from moderate to mild disability) are calculated from epidemiological or clinical data.

The model will then calculate the average cumulative costs and effects, e.g. the number of QALYs, over a defined time for an untreated and a treated cohort, and compare the groups to estimate the incremental cost (treatment costs minus cost reductions due to treatment) per QALY gained with the treatment compared with no treatment.

The time period covered by a model is divided into equal increments, referred to as "cycles". The length of the cycle is chosen to represent a clinically meaningful time interval. For instance, weekly cycles in a model to calculate the effectiveness of an anti-hypertensive treatment to avoid strokes would clearly be too short, while yearly cycles for a treatment of infections would be too long. During each cycle a patient may make a transition from one state to another or remain in the current state. The probabilities of making a transition from one state to another during a cycle (transition probabilities) are generally calculated from epidemiological data (natural disease history) or clinical trials data (treatment effect).

Figure 3.3. Markov state transition diagram

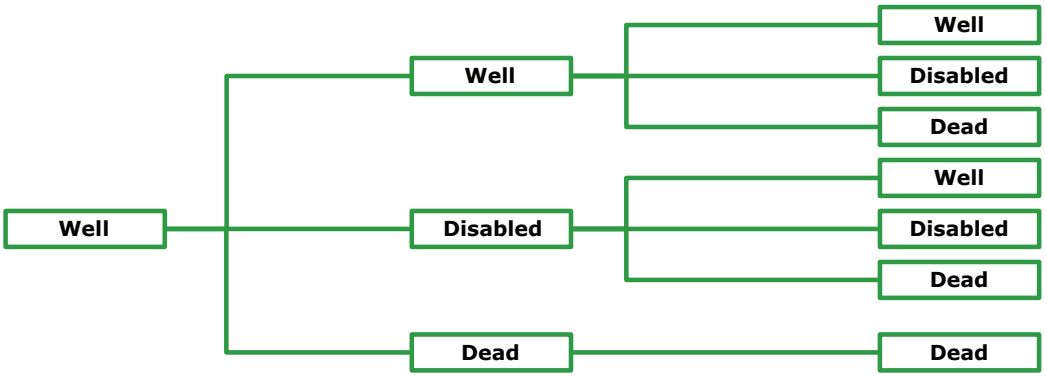


The Markov process is completely defined by how the cohort is distributed across the states at the start of the simulation and the probabilities for the individual transitions allowed during the subsequent cycles. In order for a Markov process to terminate, the model must include at least one state that the

patient cannot leave. Such states are called "absorbing states" because, after a sufficient number of cycles, the entire cohort will be included in those states. In medical examples, death is by far the most common absorbing state. Each state is assigned a utility and a cost. Cumulative utilities and costs for a given cohort are calculated at the end of the Markov process.

Simple Markov processes are often represented as so-called state transition diagrams, as shown schematically in Figure 3.3, above, or Markov cycle trees, as shown in Figure 3.4, used by a number of analytical software programmes.

Figure 3.4. Markov cycle tree



Estimating Cost-Effectiveness

Once the type of analytical technique for the economic evaluation has been defined and the input data incorporated, the expected costs and effects of the different treatment strategies can be compared.

Cost minimisation analysis

Cost-minimisation analysis is used when two or more health care interventions have the same outcomes. In such a case, the analysis can be limited to costs if conclusive evidence demonstrates that the treatments being evaluated are equally effective and that they produce no meaningful difference in health outcome. A decision maker who is responsible for all relevant costs will choose the treatment with the lowest total cost. This ensures that resources will be used efficiently.

Cost-minimisation analyses are rather infrequent, as it is rare that two treatments have identical outcomes. This is particularly the case for new treatments, as the reason for developing these is usually to improve outcomes. However, it may be the case that the trial did not show the improvement that had been expected, or that it was carried out as a non-inferiority rather than a superiority trial. In such cases, payers would perceive no additional benefit offered by the new treatment and would be interested only in funding the treatment if doing so reduces costs.

Treatment innovations, such as new formulations of existing drugs, new methods of administration, and technical improvements in procedures, can lead to fundamental changes to cost structures without affecting outcomes. For instance, many surgical interventions that traditionally were performed as inpatient procedures now can be done on an outpatient basis because of improvements in anaesthetic and surgical techniques. Frequently, the outcome of surgery is identical whether performed on an inpatient or outpatient basis. However, hospitalisation costs differ significantly, so the two alternatives can be compared on a cost-minimisation basis.

The choice between different modes of administration (e.g. oral administration, intravenous infusion, intramuscular or subcutaneous injection) is another example of a situation where costs will differ but outcomes may be similar or identical. Total costs will be influenced by considerations such as the duration of the infusion, or the number of injections required in a given period of time, and whether a trained health care professional must assist. An illustration of the cost-minimisation analysis technique is shown in Box 3.3.

Box 3.3. Illustration of cost minimisation analysis

Medicine A and medicine B both reduce one-year mortality from 25% to 15%, at a price of €10,000 and €20,000, respectively.

Medicine A requires careful dose titration in an inpatient setting and monthly laboratory tests, whereas medicine B is taken orally and requires yearly laboratory testing. Additional hospitalisation and laboratory costs for medicine A are estimated at €12,000 and those for medicine B at €500.

Total costs for medicine A are €10,000 + €12,000 = €22,000

Total costs for medicine B are €20,000 + €500 = €20,500

Medicine B, despite its acquisition cost being double that of alternative A, reduces total costs by €1,500. Given that identical outcomes are achieved by both medicines, medicine B is the cost-minimising option.

Study Example 3. Cost-minimisation analysis—kidney transplantation

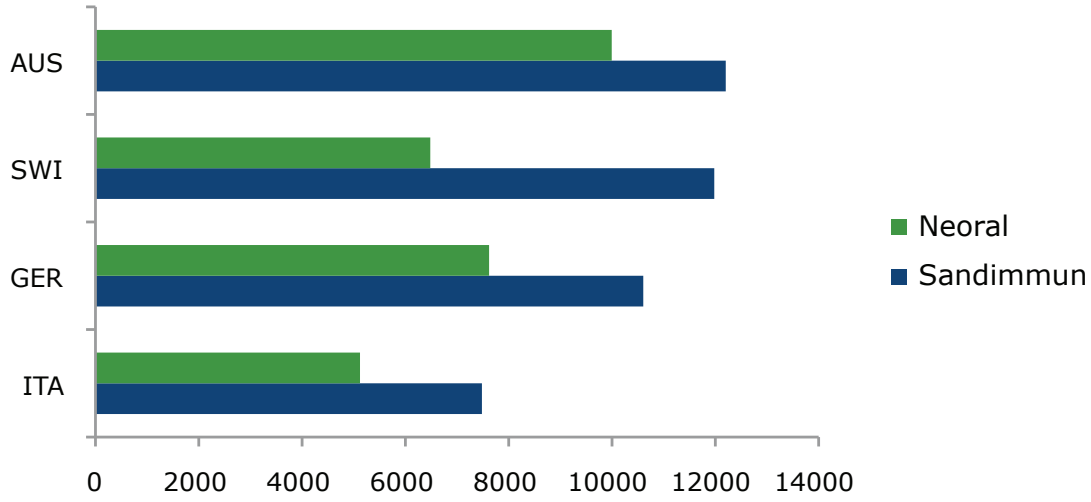
This example illustrates a within-trial analysis in a field where standard treatment protocols exist and clinical trials generally conform to these protocols. It also illustrates a situation where a cost-minimisation analysis was planned, but results rendered the interpretation somewhat difficult.

Transplantation in the treatment of chronic renal failure has been shown to be less costly and more effective in the long term than dialysis. The addition of cyclosporin A to the immunosuppressive regimen was shown in several analyses to improve short- and long-term graft survival and to reduce total costs of transplantation by reducing hospitalisation costs. The pharmacokinetic profile of cyclosporin A, however, required close monitoring of plasma levels in order to maintain immunosuppression at an optimal level and to avoid either costly episodes of graft rejection at low plasma levels or adverse effects of the medicine at high plasma levels. An improved formulation of cyclosporin A was shown to have a linear and predictable pharmacokinetic profile, and plasma levels within the optimal therapeutic window were easier to achieve. This was expected to improve the efficiency of performing transplantations due to reduced need for monitoring.

Resource use in a three-month, double-blind clinical trial in *de novo* transplant patients in four countries that compared the two galenical forms of cyclosporin A was analysed retrospectively. The clinical outcome in this trial—graft survival at three months—was expected to be identical for both groups and the trial was powered for no difference. A cost-minimisation analysis was therefore undertaken (Hardens et al, 1995). The perspective of the analysis was that of a hospital and only direct hospital costs were included. A cost advantage was shown for the new formulation, with the savings mainly due to reduced need for monitoring rejection episodes, i.e. fewer concomitant immunosuppressive medicines and shorter hospitalisations, as illustrated in Figure 3.5.

From a strictly methodological point of view, the results of this analysis must be considered as indicative only. The number of patients in the trial was small and the study was not powered to show a difference in rejection episodes (which indeed explained most of the difference in the costs). Observational studies to assess cost savings over a longer period of time and in larger samples would be required to confirm these findings.

Figure 3.5. Retrospective cost analysis of managing *de novo* transplant patients (1994 CHF)



Source: Hardens et al (1995)

Cost-Effectiveness Analysis

The average cost-effectiveness ratio represents the cost, on average, of achieving a specific outcome with a given treatment—for instance, the cost of saving one life year (as illustrated in Box 3.4). As this involves no explicit comparison, one has to assume a hypothetical and quite unrealistic scenario where the alternative to that treatment involves no costs and no effects. While average cost-effectiveness may be useful in terms of indicating a general cost level, it provides no relevant information for decisions about the allocation of resources. In most cases, the relevant choice is whether or not to replace an existing treatment with another that is more effective, but also more expensive. In such cases, an estimate is needed of the additional resources that must to be spent to obtain the additional benefit. The relevant measure in economic evaluation is therefore the ICER, which indicates the cost of producing one extra unit of benefit—a life year saved, for example, as illustrated in Box 3.4.

Box 3.4 Illustration of incremental cost-effectiveness analysis

Treatment A is the standard treatment and treatment B is a new therapy.

Treatment A reduces one-year mortality from 25% to 15% at a cost of €1,500.

Treatment B reduces one-year mortality from 25% to 10% at a cost of €2,000.

Treatment A will thus save ten life years per 100 patients and B will save 15 life years per 100 patients.

Average cost-effectiveness ratios

A: €15,000 per life year saved (€1,500 / 0.10)

B: €13,333 per life year saved (€2,000 / 0.15)

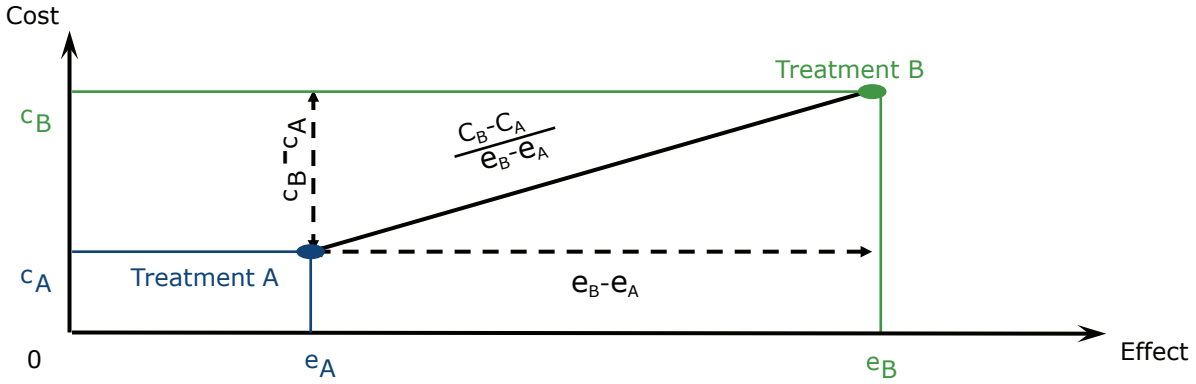
ICER of B compared to A:

€10,000 per additional life year saved ((2,000 - 1,500) / (0.15 - 0.10))

The ICER is calculated by dividing the difference in cost of two treatments by the difference in their effects (as illustrated in Figure 3.6). If a treatment is both more effective and less costly, it is the "dominant" alternative and the decision is straightforward. When the choice is between treatments where one is more effective, but also more costly, the relevant information for making the decision is the additional cost of achieving the additional outcome. The decision maker can then decide whether

or not to choose the more costly option based on a consideration of whether the extra cost is justified by the additional benefit obtained.

Figure 3.6. Incremental cost-effectiveness ratio



Study Exampe 4. Incremental cost-effectiveness analysis—onychomycosis

An example for onychomycosis (fungal infection of toenails and fingernails), although not recent, highlights a number of issues that need to be considered when performing ICER analyses. These include:

1. How to build a decision tree that reflects relevant information regarding the disease, treatment and management of care

2. How to use both data from the clinical trial as well as other data

3. How to construct a theoretical treatment pathway for a new product using an expert panel.

The methodology has evolved, however, since this analysis was performed. First, as discussed in chapter 2, "disease-free days" is not a particularly useful outcome measure for budget allocation purposes. Decision makers in most countries prefer a generic measure such as the QALY. Second, using physician panels to provide input data is no longer acceptable in most circumstances. Treatment modalities can be assessed in observational studies, as can treatment of adverse events. However, practical use of a new drug cannot be observed and panels can be used to provide input on expected use. Similarly, if a new treatment has a very specific adverse event profile not seen with current treatments, the incidence of these specific adverse events must be estimated from the clinical trial. Their treatment, if not collected in the trial, is discussed with an expert panel.

The study discussed here was performed when a new oral treatment for onychomycoses of the toenail was to be introduced in a market where three other oral products were used. Arikian et al (1994) conducted cost-effectiveness analyses of the four oral medicines in 13 countries. As an example, the analysis for treatment of toenail infections in Austria is summarised here, from the perspective of the payer, i.e. the national health insurance fund.

The medicines compared were griseofulvin, itraconazole, ketoconazole and terbinafine. Treatment modalities with these drugs as primary therapy were established in each country with a group of dermatologists. The incidence of adverse effects for each drug was established using a meta-analysis of published data and the treatment of these adverse effects was discussed with practitioners. The cost of one course of therapy was then calculated. The results are set out in Table 3.1.

Table 3.1 indicates that total treatment costs for the new therapy, terbinafine, were lower than for the comparators. If effectiveness did not differ across the treatment options, this simple cost-minimisation analysis could be sufficient to recommend the use of terbinafine. However, effectiveness needed to be estimated, as the relevant comparators were not included in the placebo-controlled clinical trials.

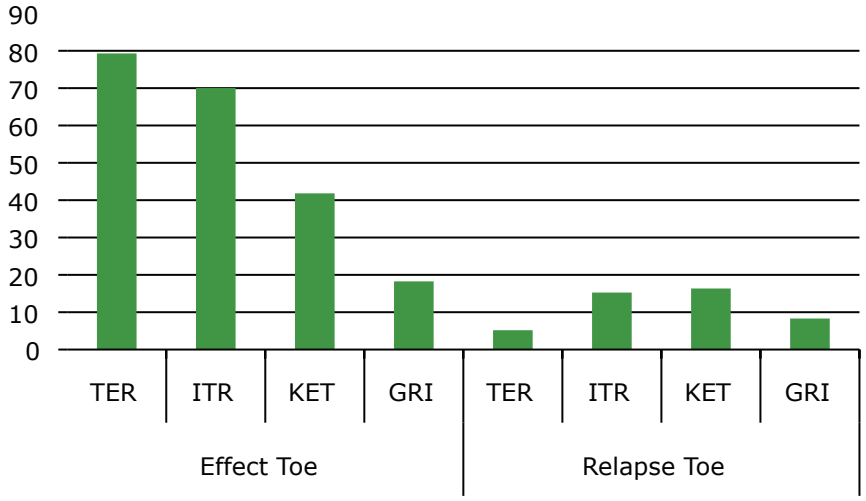
A clinical decision tree for patient management over a two-year time frame was therefore elaborated with teams of dermatologists in the different countries. Clinical outcomes were established through a meta-analysis of published clinical data, and probabilities for success, failure and relapse were estimated (Figure 3.7). These were then incorporated into the decision tree (Figure 3.8), enabling the analysis to be extended beyond the clinical trial to two years.

Table 3.1. Treatment of fungal infection of the toenail with four oral drugs: cost of one course of therapy in Austria (in ATS)

Cost item	Griseofulvin	Itraconazole	Ketoconazole	Terbinafine
Drug acquisition	3,990	12,240	6,240	9,408
Medical consultation	1,100	440	880	440
Lab tests required	9,466	2,641	9,466	2,641
Treatment of side-effects	10	34	694	4
Total costs	*14,566*	*15,355*	*17,280*	*12,493*

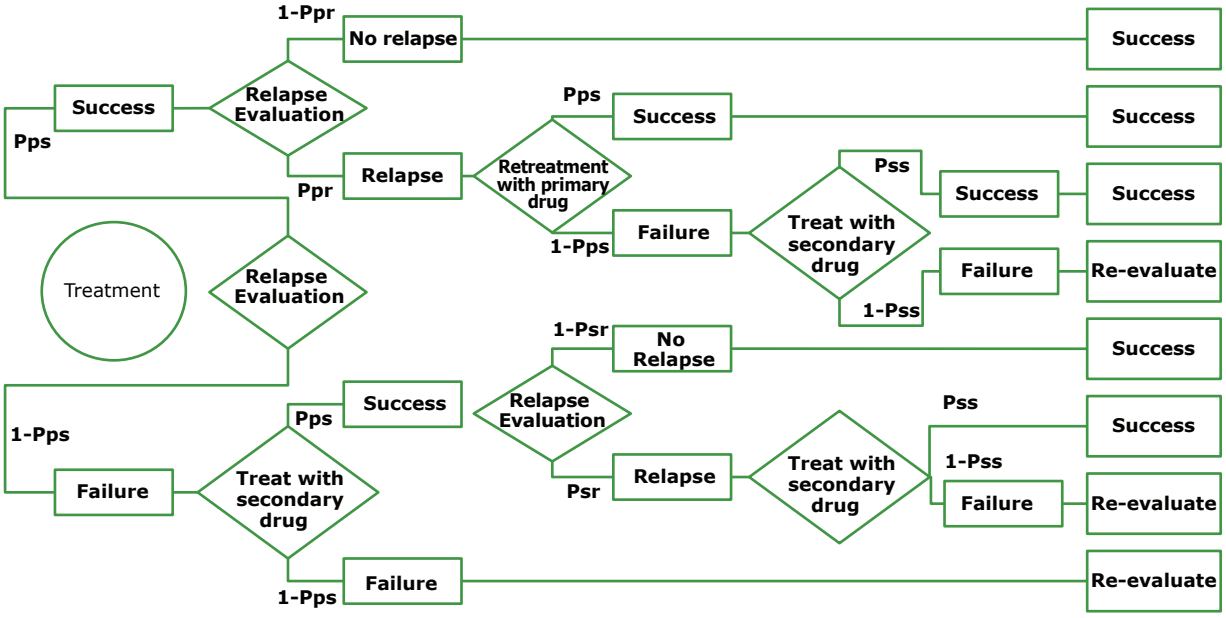
Source: Adapted from Arikian et al (1994)

Figure 3.7. Effectiveness of four treatments for onychomycosis: proportion of patients achieving cure and experiencing reinfection after one course of treatment



Source: Arikian et al (1994)

Figure 3.8. Clinical decision tree for treating fungal infection of the toenail (two years)



Key: Pps = probability of success with primary treatment; Pss = probability of success with secondary treatment; Ppr = probability of relapse with primary treatment; Psr = probability of relapse with secondary treatment
Source: Arikian et al (1994)

Costs for each individual activity in the decision tree (treatment, evaluation, testing) were established for each country and multiplied by the probabilities for each branch to calculate the average cost per treated patient for each of the four drugs. Table 3.2 shows the calculation for Austria. The expected cost of each branch is calculated by multiplying the cumulative cost of all activities in each branch by the probability of a patient following the respective branch. As the new treatment appeared to be more effective, the difference of cost compared to the comparators increased and this analysis might have provided a sufficiently compelling argument in jurisdictions that do not request cost utility analysis.

Table 3.2 Average cost per treated patient over two years (in ATS)

Decision branch	Griseofulvin	Itraconazole	Ketoconazole	Terbinafine
Branch 1	1,529	7,701	4,371	9,136
Branch 2	1,835	5,127	4,822	1,162
Branch 3	228	627	571	143
Branch 4	63	174	158	65
Branch 5	16,326	6,374	12,890	3,031
Branch 6	1,518	587	1,164	1,565
Branch 7	169	65	129	174
Branch 8	4,844	1,892	3,825	1,892
Total	*26,512*	*22,547*	*27,930*	*17,167*

Source: Adapted from Arikian et al (1994)

The measure of effectiveness was defined as "disease-free days". Using this measure, the least costly treatment (terfenadine) was also found to be the most effective. It therefore dominated the other alternatives, eliminating the need for an incremental cost-effectiveness analysis.

Study Exampe 5. Markov models in cost-effectiveness analysis—glaucoma

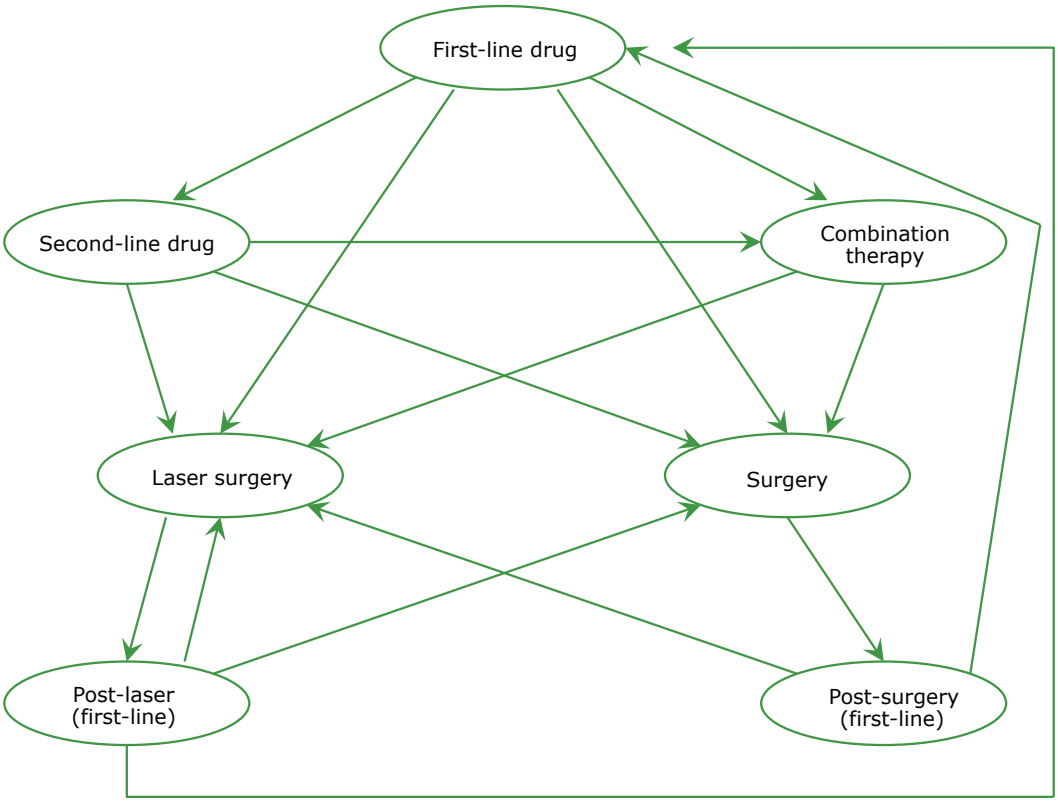
This study example refers to an early study using a Markov model with a relatively limited number of states and a short time frame. It provides a simple illustration of how Markov models can be built and used in a flexible way and how data from different sources can be combined in the absence of comparative clinical trials.

In this case, the two new treatments each had been compared to standard therapy at the time, but not to each other. Currently, the evidence from all available trials would be incorporated in an indirect comparison to overcome the lack of the required direct comparative evidence. However, not all jurisdictions accept indirect evidence, but instead require direct comparative data from the same trials that provide the clinical evidence.

In the 1990s, several new topical treatments for glaucoma were introduced. These treatments targeted elevated intra-ocular pressure (IOP), which is considered the major risk factor in glaucoma. The new treatments were more expensive than the established alternatives, but they also were more effective in lowering IOP. An economic evaluation to determine whether the additional investment could be justified was therefore required. However, the absence of a clear link between the risk factor (IOP) and the final outcome (blindness) made a cost-effectiveness analysis difficult. No epidemiological data were available that permitted calculation of the annual risk of developing blindness at given levels of IOP, controlling for age and ocular co-morbidity.

As an alternative solution, this study estimated the cost of different treatment strategies over two years, given the clinical effectiveness of these strategies in controlling IOP (Kobelt and Jönsson, 1999). The consequence of treatment was hence incorporated indirectly by using clinical data to calculate the proportion of patients who achieved and maintained IOP levels below the desired clinical target level, thus reducing the need for intensive management (treatment changes, surgical interventions). Treatment strategies were represented in a Markov model (Figure 3.9), combining clinical trial results and observational data (as presented in Study Example 2).

Figure 3.9. Structure of the Markov model

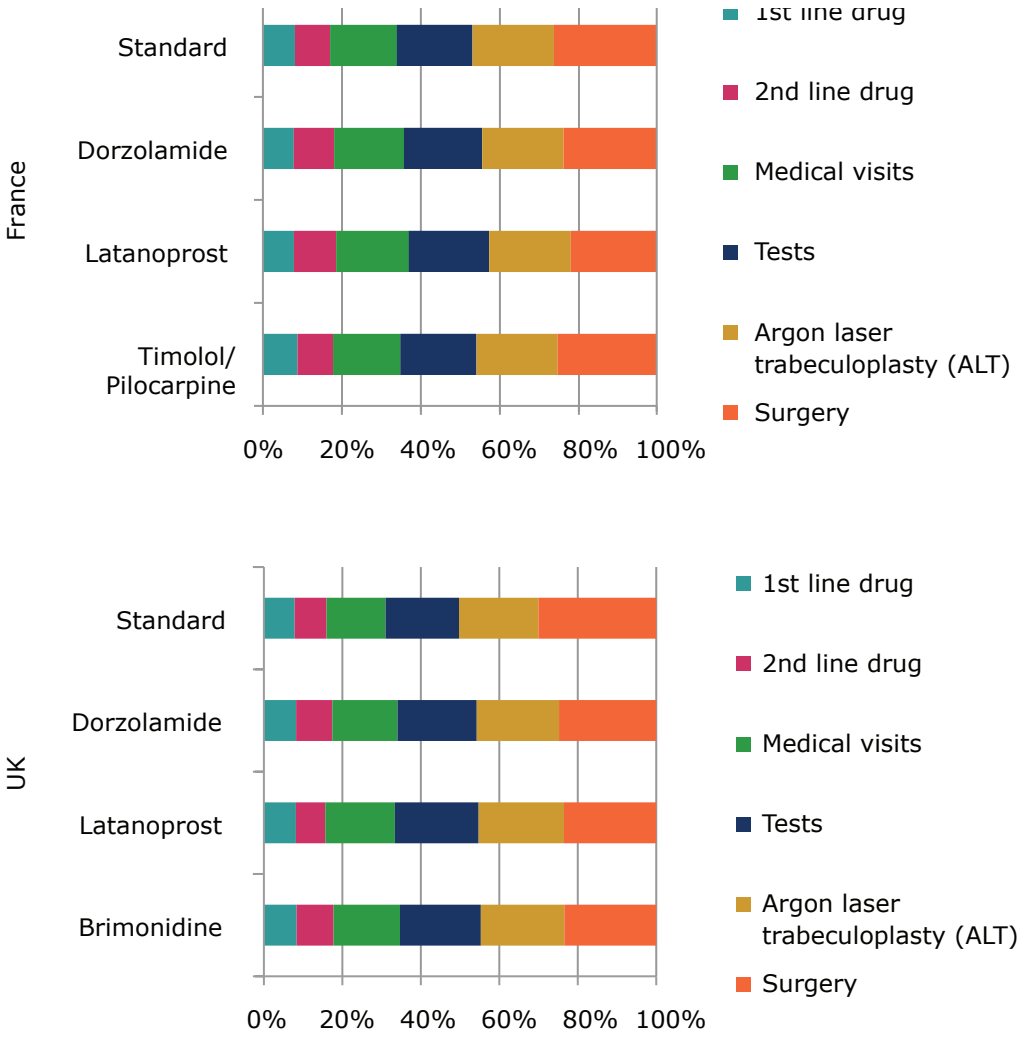


Source: Kobelt and Jönsson (1999)

In this Markov model, states are defined by the treatment patients receive rather than by clinical measures, as is more typical. The cycle length is one month, to account for short treatments such as surgical interventions and follow-up. However, treatment changes can take place only during consultations, which occur on average every three months, as determined from an observational study. Technically, patients who fail treatment at one or two cycles are placed in a holding state (called a "tunnel state") until they switch treatment after three months. Treatment switches observed in that study were used to calculate transition probabilities for current treatment. The effect of the new treatments, which were approved as second-line treatments, was then calculated by replacing the established second-line drugs with the new drugs in the model. The proportion of "controlled IOP patients" at every three-month interval were calculated from clinical trials and used as the new transition probabilities. The cost per cycle was based entirely on resource use in the observational study. In the absence of long-term data for any of the agents compared, the model duration was limited to the first year after diagnosis.

The model found that, with the new drugs, the distribution of costs across categories of resources changed (Figure 3.10). Better control of IOP reduced the need for surgery and resources were hence shifted from surgery and hospitalisation to drug treatment. The net result was that costs did not increase and, in countries with a high rate of surgery, had the potential to decrease for some of the second-line treatments, as shown for France and the UK in Table 3.3.

Figure 3.10. Distribution of costs by category of resource use: France and the UK as examples.



Source: Kobelt and Jönsson (1999)

Table 3.3. Average cost per patient with different second-line treatments during the first year after diagnosis for intervention

	First-line drug	Second-line drug	Visits	Tests	Argon laser trabeculo-plasty	Surgery	Total one year
France (FF)							
Standard strategy	739	68	750	211	129	492	2,389
2nd line: Dorzolamide	814	179	751	198	74	289	2,305
2nd line: Latanoprost	753	272	754	186	25	97	2,087
2nd line: Timolo/pilocarpine	811	11	751	206	107	419	2,305
UK (GBP)							
Standard strategy	69	1	63	33	12	87	265
Dorzolamide	70	8	63	31	5	33	210
Latanoprost	67	15	63	30	2	15	192
Brimonidine	70	9	63	31	5	33	211

Source: Kobelt and Jönsson (1999)

Study Example 6. Combining data in cost-effectiveness analysis—chronic heart failure

A study by Cleland et al (2001) illustrates how short term data from a clinical trial can be combined with epidemiological data to estimate life expectancy and demonstrates a simple method that can be used when no detailed epidemiological data are available. The example also illustrates the use of two techniques, discounting and estimating costs in added years of life, which are further explained later on.

The prognosis for patients with chronic heart failure (CHF) is poor, with five-year mortality estimated at 62% for men and 48% for women. Although preventive treatment with ACE-inhibitors has improved survival, one-year mortality in patients with very severe disease (New York Heart Association [NYHA] class IV) remains around 50%. The goal of treatment in CHF is to improve survival without increasing morbidity and related hospital inpatient costs, which constitute 60–75% of all costs.

In a one-month clinical trial, a new inotropic agent, levosimendan, was shown to improve survival significantly in patients with severe CHD (NYHA III–IV) compared with the standard treatment, dobutamine. The study included 199 patients admitted to cardiac care units in ten European countries who received a 24-hour infusion of either of the drugs and were followed for one month. Survival was followed up for six months, but resource utilisation data were only available for the first month. Hospitalisation data were therefore collected retrospectively for the remaining five-month period and, probably due to the severity of the disease, it was possible to obtain data for 99% of the patients.

Resource utilisation in the economic analysis included study drugs and inpatient care. Concomitant medication and outpatient visits were omitted from the analysis, as there was no difference in any of the types of drugs between the two groups and the number of visits was protocol-driven. For each country, the costs per inpatient day in a cardiology ward, in a cardiac care unit, and in intensive care, as well as the list price per mg for dobutamine, were obtained. These costs were converted into euros and a mean cost was calculated. Because the price for levosimendan was not yet available, an expected price was assumed for the analysis.

In the new treatment arm, 75 of 102 (73.5%) patients were alive at six months, compared to 61 of 97 (62.9%) patients in the standard treatment arm. Mean survival over six months was 157 days and 139 days, respectively. The risk of death with the new treatment was thus reduced by 32% in relative terms and 10.6% in absolute terms. Inpatient costs were similar, so the difference between the two groups was almost entirely due to the study drug (Table 3.4).

Table 3.4. Resource utilisation and cost over six months (in EUR)

	Dobutamine Mean (SD)	Levosimendan Mean (SD)	Difference
Life years gained (undiscounted)	2.268	2.636	0.369
Life years gained (discounted 3%)	2.133	2.479	0.346

Source: Cleland et al (2001)

The difference in the risk of death at the end of the trial, however, does not capture the full benefit, as the two groups differed in terms of the proportion of patients that lived on beyond the clinical trial. It is thus necessary to extrapolate the gain in life expectancy beyond the trial. This requires good epidemiological data for a similar patient group, as survival in CHD depends on age and the severity of the disease. In this case, patient-level data from a longitudinal study were available: in the 1987 CONSENSUS trial, patients with severe CHD were randomised to an ACE-inhibitor or placebo and followed until death (Swedberg, Kjekshus and Snapinn, 1999). These patients were similar to the patients in the clinical study with levosimendan in terms of age and sex distribution, as well as one- and six-month mortality (9% versus 11%, 29% versus 31%, respectively). As all patients in the levosimendan trial had received ACE-inhibitors, the ten-year survival data of the treatment group in CONSENSUS were used to extrapolate life expectancy.

In CONSENSUS, patients were randomised to placebo or ACE-inhibitors for six months, after which all patients still alive continued on ACE-inhibitors for ten years. At the last follow up, five patients remained alive. Survival was estimated for the period after the double-blind phase to match the levosimendan trial. Mean survival for patients alive at the end of the double blind period was 941 days or 2.6 years (assuming that the five patients still alive die immediately). This "conditional" survival should therefore be interpreted as the lower limit and actual survival is expected to be somewhat higher. Thus, the cost-effectiveness analysis assumed a life expectancy of three years, discounted at 3% per annum, at the end of the levosimendan trial. Matching survival in the two datasets, the gain in life expectancy with levosimendan is 0.369 years as shown in Table 3.5.

Table 3.5. Gain in life expectancy using epidemiological data

	Dobutamine Mean (SD)	Levosimendan Mean (SD)	Difference
Life years gained (undiscounted)	2.268	2.636	0.369
Life years gained (discounted 3%)	2.133	2.479	0.346

Source: Cleland et al (2001)

If no patient-level data are available, as in this case, the calculation could be done in a simplified way. Based on the clinical trial, for every 100 patients treated in each group, 73.5 would be alive in the levosimendan group and 62.9 in the dobutamine group at six months. The difference in survival is hence 0.0265 x (0.5 x [0.5 years x 10.6]). If mean life expectancy after the trial is three years and assuming that survival is linear, the survival gain after the trial is 0.318 x (0.5 x [6 years x 10.6]). The total is then 0.345 years (undiscounted). If the mean life expectancy is only two years, the gain would be 0.239 years on average. In this case, the two methods yield similar results indicating that survival in CONSENSUS was probably near-linear. This is not necessarily the case; the two groups in this trial could have developed differently. Cost-effectiveness, then, is calculated as the incremental cost (1,154 euro) divided by the incremental benefit (0.346 years) as shown in Table 3.6.

Table 3.6. Cost per life-year gained

	Conditional mean survival (years)	Cost per life-year gained (EUR)
Base case	3	3,335
Sensitivity	2	4,603
Sensitivity	4	2,639

Source: Cleland et al (2001)

CHF is a typical example of a disease for which costs incurred during added years of life may play a role. These costs were calculated for Sweden in this example. Where relevant to the study and indication, Swedish reimbursement authorities request such data because it may affect the cost-effectiveness ratio. Using published data on production and consumption for different age groups to estimate future costs (Johannesson, Meltzer and O'Conor, 1997), the cost per life-year gained increased from 27,700 SEK (€3,080) to 190,000 SEK (€21,000). More details on costs in added years of life appear below.

Technical Issues

The onychomycosis and glaucoma study examples cover relatively short time frames, and data on both costs and clinical effects were directly available from the clinical trials from other datasets. The CHF example covered a longer time frame and involved extrapolation of data beyond the clinical trial. It highlights a number of technical issues associated with cost-effectiveness analysis.

Time perspective or discounting

Many economic analyses cover several years, and costs and effects often do not occur at the same time. In order to compare treatments or expenditures and benefits in different time periods, discounting should be used (see Box 3.5). Discounting can be applied to all forms of economic analysis and is requested by all guidelines. However, while the discounting of costs is uncontroversial, whether benefits also should be discounted is still a matter of debate. The argument against the use of discounting is that the value of a benefit does not depend on whether it occurs now or at a later time. The general rule is therefore to present results for both costs and benefits in the "undiscounted" and discounted form using a common rate for both.

Box 3.5. Discounting

Discounting is a technique that allows comparison of costs and benefits that occur at different times. This is particularly important in health care where costs often occur immediately, but benefits may occur at a later stage, e.g. with preventive programmes such as vaccination, lipid-lowering and anti-hypertensive therapy, or where treatment continues over a long period, e.g. in the long-term treatment of chronic and progressive illnesses.

Discounting is not a correction for inflation. Rather, it reflects time preference (the desire to have benefits earlier rather than later) and the opportunity cost of capital, i.e. the returns that could be gained if the resources were invested elsewhere.

The technique is straightforward. For example, based on a discount rate of 5%, a cost of €1,000 occurring in one year's time is considered to be worth only about 95% at present value, i.e. approximately €950. €1,000 in two years would be worth €907 today; the same amount in three years would be €864, and so forth. Alternatively, €864 invested at 5% will grow to €1,000 in three years' time, and €907 will grow to €1,000 in two years' time. The adjustment that has to be made to future flows to express them in present values is:

$$\frac{1}{(1+i)^t}$$

where *i* is the discount factor and *t* is the number of years.

Thus, €1,000 in five years at a discount rate of 3% is worth €863 (1,000 / [1 + 0.03)5]; €1,000 in ten years has a value of €744 today.

Cost of added years of life

A further technical issue illustrated in the CHF example is the fact that a patient with a life-threatening disease whose life is saved or extended with treatment will continue to use health care resources in the added years of life. Techniques and issues for adding such "future costs" are described in Box 3.6.

Box 3.6 Future costs

In general, only the costs related to the specific disease have been included in cost-effectiveness analyses, which has led to criticism that this overestimates the extent of cost effectiveness. Some have argued that all future related and unrelated costs should be included in the analysis (Meltzer, 1997). However, to include all health care consumption in the added years of life is not without problems, since it is not entirely obvious that the average consumption can be applied to all age groups. While overall expenditures increase as people live longer, expenditure per year as well as expenditures in the last year of life decrease with increasing age. The following table illustrates this, although this distribution of costs can be expected to have changed somewhat in the years since the study's publication.

Table 3.7a. Lifetime health care costs for people living to 65 or older in the US (Medicare data, USD)

Age at death	Lifetime health care costs	Cost per extra year of life	Cost during the last two years of life
65-70	13,000	0	0
71-80	35,500	3,600	23,000
81-90	56,000	1,200	21,000
91-100	63,000	400	15,000
>100	66,000		8,000

Source: Lubitz, Beebe and Baker (1995)

Also, to make the analysis complete, general consumption (not including health care) and production also should be included. In younger age groups, production is greater than consumption, but the opposite is true after retirement age.

Table 3.7b. Annual consumption and production in different age groups in Sweden (1995 SEK)

Age groups	Consumption minus production		Production	Consumption
	Private	Public		
35-49	98,000	32,000	214,000	-84,000
50-64	113,000	32,000	182,000	-37,000
65+	77,000	82,000	0	159,000

Source: Johannesson, Meltzer and O'Conor (1997)

Shown below are the results of a cost-effectiveness analysis in hypertension stratified by age and risk that included costs in added years of life. The effect on the cost-per-life-year ratios in the younger age groups is minimal, whereas the ratios change considerably in the older age groups.

Table 3.7c. Cost per life-year gained with and without costs in added years of life through treatment of hypertension in Sweden (1995 SEK 000's, 3% discount rate)

Diastolic blood pressure mmHg	<45 years		45-69 years		>70 years	
	Men	Women	Men	Women	Men	Women
90-94						
• **without**	818	1,825	58	153	29	22
• **with**	825	1,869	161	263	190	182
95-99						
• **without**	679	1,394	29	876	15	7
• **with**	686	1,423	131	204	175	168
100-104						
• **without**	562	1,022	7	36	7	0
• **with**	569	1,051	95	139	168	161

Source: Johannesson, Meltzer and O'Conor (1997)

Patient groups, stratification of risk and sub-analysis

Cost-effectiveness analysis may encounter situations where a treatment may be very cost-effective in one patient group, but not at all in another. When clinical trials are large, patients can be stratified by specific level of risk, as illustrated in Box 3.6, above. The cost per life-year saved is high in younger patients with a low relative risk of a fatal cardiac event (below 45 years of age, 90-94 mmHg diastolic blood pressure). This is because few events will be avoided in absolute terms; the ratio will decrease as the risk increases to over 100 mmHg. The same pattern can be seen in patients aged over 70 years, but at any given level of blood pressure. The risk in this group is much higher and hence the cost per life-year saved is substantially lower. As is well known in cardiology, women overall have a lower risk of a cardiac event, which translates directly into higher cost-effectiveness ratios.

In recent years, payers have increasingly restricted the funding of drugs to defined groups of patients where the risk is highest or the unmet need is greatest. This is particularly true for drugs with a higher price or with serious adverse events. It is not always possible to plan for this in the clinical development process, however, and subgroup analysis therefore often lacks statistical power, resulting in high levels of uncertainty. As a consequence, pricing and reimbursement decisions are often associated with requests for follow-up studies after marketing commences.

An instructive example is a comparison of the treatment of MS with natalizumab and fingolimod. In 2006, natalizumab was found to have a low risk of a particular lethal adverse event and was authorised as a second-line treatment for patients with highly active forms of the disease, despite the fact that second-line use had not been studied in the clinical trial. In a similar fashion, fingolimod was developed as a first-line treatment. However, its adverse event profile prompted the European regulatory agency in 2011 to limit treatment in Europe to the same patients as natalizumab, despite the absence of statistically significant data in that group and a relevant comparator. Reimbursement authorities in a number of countries then asked for "real world" follow-up studies. In both cases, regulators appeared to be persuaded of the products' efficacy and were willing to accept the uncertainty of effectiveness in second-line therapy, provided that they received subsequent proof of clinical effectiveness and cost-effectiveness.

Uncertainty: sensitivity analysis and confidence intervals

Economic evaluations must rely on different sets of data and on modelling in order, for instance, to link short- or medium-term clinical effects (e.g. lowering of hypertension) to long term outcome (e.g. avoidance of stroke) or to incorporate resource utilisation. Assumptions often must be made because of the uncertainty in both clinical and resource utilisation data. As the credibility of the results will depend on the quality of the data used, it is important to explore the effect of making alternative assumptions and to perform extensive sensitivity analyses, particularly for those parameters with the highest degree of uncertainty. This is discussed in Box 3.7.

Box 3.7. Sensitivity analysis

A sensitivity analysis examines the effect on the study results of systematic changes in key assumptions or parameters. For example, what would the impact be on the results if the effectiveness of a treatment were increased or decreased; or if the costs of any of the resources used were doubled or halved; or if the incidences of side effects were lowered or increased? Sensitivity analysis helps explore some of the uncertainty related to potential variability in the basic data and the sample population, and to extrapolate from one setting to another. It helps identify which parameters or assumptions have the greatest effect on the outcome and the stability of the results.

Sensitivity analysis, in its simplest form, involves varying one or more parameters across a possible range. Other variations include finding the threshold value of a variable above or below which the conclusion of the study will change, and analysing the impact of assuming extreme values of a variable.

When data are collected in the context of a trial, the observed variance in the data allows statistical techniques to be applied. Generally, the data are highly variable, particularly cost data, and it is becoming standard practice to present confidence intervals for incremental cost-effectiveness ratios. Because the range of estimates for effects can come close to, or sometimes overlap zero, the corresponding range for the cost-effectiveness ratio can approach infinity and the confidence interval can be extremely wide. One approach that avoids this problem, and which arguably gives a more meaningful measure of the variation in the data, is to calculate a cost-effectiveness acceptability curve, as shown in Box 3.8

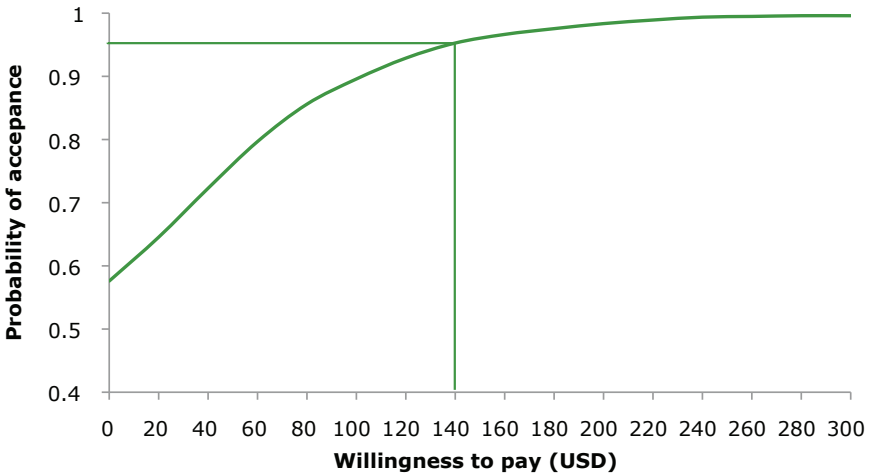
Box 3.8. Confidence intervals and cost-effectiveness acceptability curves

The variability in cost and effect data is often high, leading to ICERs with a high uncertainty. This variability can be represented statistically in the form of a confidence interval around the ICER. However, because of the properties of the ratio statistic, this is not a straightforward matter.

Different approaches have been developed to estimate confidence intervals for ICERs. Those being used currently include the confidence box method (Wakker and Klaassen, 1995), Taylor approximation (delta method, O'Brien et al, 1994), Fieller's interval method (Fieller, 1954), and bootstrap methods (Briggs, Wonderling and Mooney, 1997). In addition to the possibility that confidence intervals can be extremely wide, their interpretation is complicated by the presence of negative ratios within the interval, since these can have one of two diametrically opposed meanings. First, they can imply that the treatment of interest is more effective and less costly than (i.e. dominates) the comparator for some values of costs and outcomes. Second, they can imply that it is more costly and less effective than (i.e. is dominated by) the comparator for some values of costs and outcomes.

A potentially more meaningful way of presenting the same data is the cost-effectiveness acceptability curve, an example of which is shown in Figure 3.11. For a range of values of willingness to pay for health benefits, P, the cost effectiveness curve shows what proportion of the corresponding estimates of the ICER are acceptable. Using data on costs and outcomes, the probability that the ICER falls below the required limit can be derived if an assumption is made about the distributions of mean costs and mean outcomes. Van Hout et al (1994) present an illustration assuming normal distributions.

Figure 3.11. Cost-effectiveness acceptability curve



Alternatively, the distribution of the ICER may be estimated by "bootstrapping" from the observed samples. In this approach, bootstrap samples are generated by selecting patients one at a time from the observed data, with replacement, until a sample the same size as in the original data is obtained. In general, some patients' data will be selected more than once and others not at all. Taking the mean of costs and outcomes for this sample gives one bootstrap estimate of mean costs and mean outcomes. The distributions of mean costs and mean outcomes, and thus of the ICER, are then estimated by repeating this procedure a large number of times, perhaps several thousand. For a more detailed discussion of the method, see Briggs, Wonderling and Mooney (1997).

Cost-effectiveness acceptability curves also can be estimated when costs and outcomes are simulated by a modelling process, for example, in a Markov model. A distribution for the ICER can be derived using probabilistic sensitivity analysis in which the parameters of the model are assigned distributions and the model is run a large number of times, with re-sampling from these distributions on each occasion. A detailed illustration of the application of this method is provided by Briggs (2000).

The acceptable range of estimates for the ICER clearly will include all those cases where the treatment of interest dominates the alternative and those cases where the additional cost per unit of health outcomes is no greater than P. In other words, a new treatment, A, should replace the old treatment, B, if the ICER, (costsA-costsB)/(effectsA-effectsB), is less than or equal to R. The proportion of estimates falling within the threshold P is often interpreted, from a Bayesian perspective, as the probability that the intervention is cost-effective.

A statistic that avoids some of the problems of ratios is the net benefit (NB), defined as the monetary value of incremental effects less incremental costs. The net benefit therefore can be expressed as NB = P x (difference in effects – difference in costs). If it is greater than 0, the new therapy should be adopted. Plotting the proportion of estimates for which the NB is positive against different values of P gives a presentation equivalent to the CEAC.

Meta-analysis and network meta-analysis: indirect comparison

Meta-analysis long has been used in HTA. Systematic reviews combine all available information on a given product into a framework that allows conclusions about effect size to be made with more statistical power than is possible using evidence from the individual trials. In the past, trials generally measured the effect of a treatment against placebo, making meta-analysis straightforward and providing a weighted average of the effect size of the different trials.

Payers now request that new treatments be compared to relevant available alternatives, typically termed an "active comparator", rather than placebo. Such head-to-head comparisons, however, often are unavailable for a number of reasons; for example, trials for products coming to the market currently were initiated before this requirement was explicit, or several alternatives exist and the particular comparator used in a trial may not be relevant to the payer in a given country. To overcome this lack of relevant data, network meta-analyses have been developed whereby trials of different

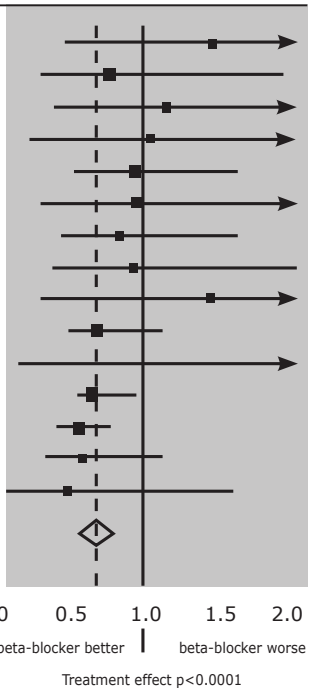
products against placebo or against each other are compared indirectly by statistical inference. While this increases the uncertainty already present in routine meta-analysis, it does provide a valuable indication of comparative effectiveness.

Box 3.9. Meta-analysis and network meta-analysis

According to the Cochrane Collaboration terminology (Higgins and Green, 2011), a meta-analysis refers to statistical methods of combining evidence, while leaving other aspects of "research synthesis" or "evidence synthesis", such as combining information from qualitative studies, for the more general context of systematic reviews. Thus, a meta-analysis combines the results of several studies that address related research hypotheses. A common measure of effect is identified and the effect size compared as a weighted average. The weighting might be related to sample sizes or patient populations within the individual studies as well as to other differences between studies. The general aim of a meta-analysis is to provide a more powerfully estimate of effect size than is possible from individual trials alone.

Figure 3.12. Meta-analysis of trials comparing beta-blockers to placebo

Study	No (%) of deaths/patients		Beta-blocker deaths	
	Beta-blocker	Control	Logrank observed-expected	Variance observed-expected
Wilcox (oxprenolol)	14/157 (8.9)	10/158 (8.9)	2.0	5.6
Norris (propanolol)	21/226 (9.3)	24/228 (9.3)	-1.4	10.2
Multicentre (propanolol)	15/100 (15.0)	12/95 (12.6)	1.2	5.8
Baber (propanolol)	28/355 (7.9)	27/365 (7.4)	0.9	12.7
Andersen (alprenolol)	61/238 (25.6)	64/242 (26.4)	-1.0	23.2
Balcon (propanolol)	14/56 (25.0)	15/58 (25.9)	-0.2	5.5
Barber (practolol)	47/221 (21.3)	53/228 (23.2)	-2.2	19.5
Wilcox ((propanolol)	36/259 (13.9)	19/129 (14.7)	-0.7	10.5
CPRG (oxprenolol)	9/177 (5.1)	5/136 (3.6)	1.1	3.3
Multicentre (practolol)	102/1,533 (6.7)	127/1,520 (8.4)	-13.0	53.0
Barber (propanolol)	10/52 (19.2)	1247 (25.5)	-1.6	4.3
BHAT (propanolol)	138/1,916 (7.2)	188/1,921 (9.8)	-24.8	74.6
Multicentre (timolol)	98/945 (10.40)	152/939 (16.2)	27.4	54.2
Hjalmarson (metoprolol)	40/698 (5.7)	62/697 *8.9)	-11.0	23.7
Wilhelmsson (alprenolol)	7/114 (6.1)	14/116 (12.1)	-3.4	4.8
Total*	640/7,047 (9.1)	784/6,879 (11.4)	-81.6	310.7



Reduction 23.1% (standard error 5.0) p<0.001
Heterogeneity between 15 trials: $\chi^2 = 13.9$; df = 14; p>0.1
* 95% confidence interval as shown for the odds ratio

Source: Lewis and Clarke (2001)

The steps in a meta-analysis are:

1. Formulation of the problem, i.e. a well-defined subject/indication
2. Search of the literature
3. Selection of studies based on study-quality criteria, e.g. the requirement for randomisation and blinding in a clinical trial
4. Decision on inclusion of unpublished studies (to avoid, e.g., publication bias)
5. Selection of dependent variables or summary measures to be allowed (e.g. differences, means)

6. Selection of the statistical model (e.g. a "fixed effect model" where the weighted average is based on study size, i.e. larger studies carry a larger weight, or a "random effect model" used to synthesise more heterogeneous research)

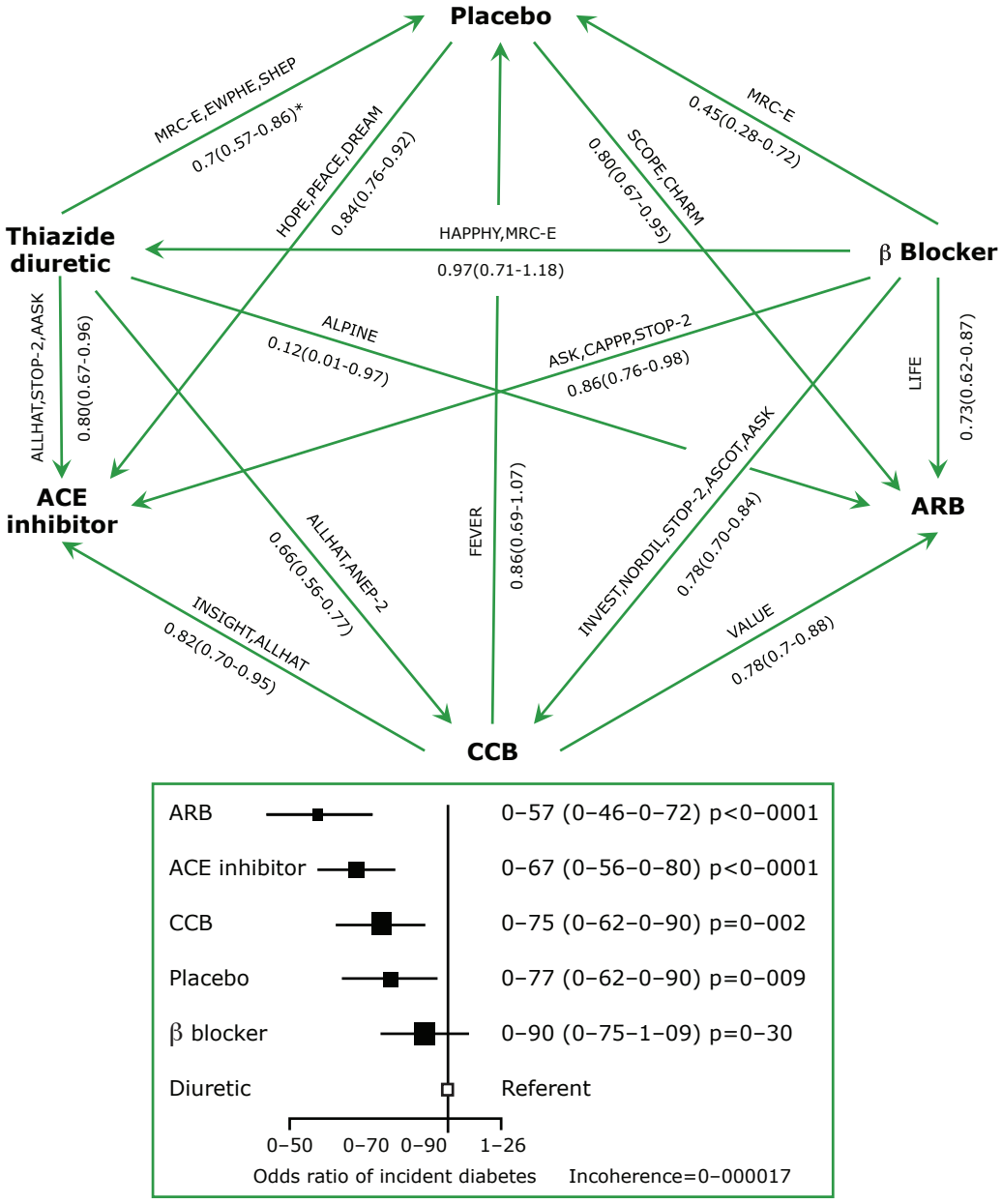
7. Transparent and complete reporting.

An example of presentation of results from a meta-analysis of trials comparing beta-blockers to placebo is shown in Figure 3.12.

Box 3.10. Indirect comparisons: network meta-analysis

Network meta-analysis is a statistical method that enables a comparison of two or more products when head-to-head data are lacking. The simplest example would be a comparison between treatment A and C in the same

Figure 3.13. Estimating the incidence of diabetes from studies of antihypertensive drugs



Source: Elliott & Meyer, 2007

patient population, where one study compares A to B and another study compares B to C. If, for instance, A is found to be superior to B, and B is found equivalent to C, then one could logically infer that A is also superior to C in this particular population. The validity of such analysis hinges clearly on the similarity of the studies included, i.e. similar studies in similar populations. The greater the number of studies included, the greater the risk of heterogeneity and uncertainty in the results.

Authorities have reacted differently to the use of indirect comparisons. While the UK and Canada explicitly accept them (Canada has issued guidelines for their use), many countries appear to remain sceptical due to the high levels of uncertainty involved.

The initial steps in network meta-analysis are the same as for classical meta-analysis, i.e. clear definition of the question (population, indication, effect measure) and literature extraction. Subsequently, odds ratios (OR) and standard errors (SE) are estimated for each comparison within studies (depending on how many arms are included) or for groups of comparable studies. For the network, the logarithm of the OR and the SE are used.

An example of such an analysis is shown below where the incidence of diabetes was estimated from 22 long-term studies of antihypertensive drugs.

Cost-Utility Analysis

Much of modern medicine is concerned with improving the quality of life, not just the duration. Hence, the effects of different health care interventions on patients' quality of life need to be considered together with survival to assess the total impact of treatment. A number of outcome measures have been designed to include both these concepts, such as the healthy-year equivalent (HYE), the disability-adjusted life year (DALY) and the quality-adjusted life year (QALY). The QALY is, by far, the most frequently used measure today. DALYs have been used almost exclusively in international comparative studies by the World Bank and the World Health Organisation and have been advanced principally as a means of estimating the overall burden of disease worldwide. The relatively onerous demands placed on respondents by the HYE method help to explain why it has been adopted only occasionally in economic evaluations.

Cost-utility analysis is a type of cost-effectiveness analysis that incorporates both quantity and quality of life by estimating the cost per QALY gained as a result of treatment. QALYs are calculated by weighting time (years of life) with a quality adjustment, called "utility", which represents the relative preference that individuals or society place on different states of health. Cost-utility analysis has two major advantages compared to other economic evaluation techniques: in addition to combining life expectancy and overall quality of life aspects, the use of a standard outcome measure allows for comparison between treatments in different disease areas that may have very different clinical outcome measures. A health care payer will need to compare different treatments to make expenditure and prioritisation decisions within its budget, across diseases and indications. It is therefore unsurprising that organisations in countries where economic evaluation is used to inform decision making prefer cost-utility analysis to other types of analysis. Economic evaluation per se does not give a value of the benefit itself, but only estimates the relative inputs required to reach a given outcome; comparison is an essential feature of resource allocation.

Having the same generic outcome, e.g. the QALY, theoretically allows the comparison of different cost-utility studies. In the past, a number of "league tables" of different health care interventions have been created, as illustrated in Table 3.8. In practice, however, such league tables should be considered with caution and are seldom used, but this does not in itself degrade the value of the QALY for decision making. Rather than comparing cost per QALY estimates explicitly, however, authorities have tended to adopt official and unofficial thresholds for their willingness to pay for a QALY gained.

Table 3.8. League table: cost per QALY for selected interventions in the UK (1990 GBP)

Treatment	Cost per QALY
Cholesterol testing and diet therapy (ages 40–69)	220
Advice to stop smoking from GP	270
Antihypertensive treatment to prevent stroke (ages 45–64)	490
Pacemaker implantation	1,100
Hip replacement	1,180
Cholesterol testing and treatment	1,480
Coronary artery bypass graft (CABG) (left main vessel disease, severe angina)	2,090
Kidney transplant	4,710
Breast cancer screening	5,780
Heart transplantation	7,840
Cholesterol testing and treatment (incrementally), all adults ages 25–39	14,150
Home haemodialysis	17,260
CABG (one vessel disease, moderate angina)	18,830
Hospital haemodialysis	21,970
Erythropoietin treatment for anaemia in dialysis patients (mortality—10%)	54,380
Erythropoietin treatment for anaemia in dialysis patients (no incremental survival)	126,290

Source: Adapted from Maynard (1991)

As is true for all study comparisons, the methodology used in cost-utility analysis is not always consistent across studies on either the input or output side. Studies may use different concepts of costs or adopt different perspectives, thus including or excluding different types of resources. They may refer to a different time horizon or to different patient sub-populations. On the output side, different studies may use different methods to estimate utilities, which can lead to differences in the values generated. Differences can originate from the way the questions to elicit preferences or utilities are asked and also from the methods themselves. The methods for eliciting utilities are explained below.

Utilities

In economic evaluation, utilities are preference weights for given health states, where a utility (score) of one generally represents full health and zero represents death. Two common types of methods are used to calculate utilities. One elicits preferences about health states directly from patients via an interview. The other uses preference-based generic quality-of-life questionnaires, such as the EuroQol Group's EQ-5D, developed in Europe and used worldwide, or the Health Utility Index (HUI) developed in Canada and used mostly in North America.

In the first approach, patients enrolled in the study are asked to rate scenarios relating to their health and the treatment they received using one of the valuation methods described below—standard gamble or time trade off. These values are then used to rate the quality of life benefits of treatment. However, it is not always possible to elicit utilities directly in clinical trials as these methods require trained interviewers, are time consuming, and difficult to integrate into the study protocol. The analyst may instead seek the values of a group of patients that is broadly representative of that to which the analysis is intended to apply.

An additional challenge associated with eliciting utilities directly from certain patient populations is cognitive ability or understanding may be impaired (e.g. in Alzheimer's disease, schizophrenia) or the disease may be so severe that it is considered unethical to seek patients' preferences (e.g. late-stage cancer). A solution to this has been to ask proxies, such as health care professionals or caregivers, to complete the valuation exercise on behalf of the patient. One must remember, however, that proxy values cannot truly replace patient values and should be avoided whenever possible.

In contrast, the preference-based quality-of-life instruments mentioned in the second approach have predefined sets of scores associated with them. These instruments are used to define a variety of "health states", each of which has a utility score that can be used to calculate QALYs directly. For example, a "tariff" has been generated for the EQ-5D for the UK that was calculated from the stated preferences of a random sample of the general public using the time-trade-off method (Dolan et al, 1995). Analysts may therefore assign a value to the health status associated with a given response to the EQ-5D questionnaire by referring to this tariff.

In collectively financed health care systems, it is argued that the general population is the most appropriate group to provide "generic" valuations that will be used to guide allocation of resources among different programmes across the health service (Gold et al, 1996). Scores collected from the general population can also avoid including the coping effects of patients (termed "habituation") into the valuation, which might reduce the value of the benefit. Patients with chronic diseases often learn to adapt to their situations and tend to rate their health states higher than a healthy person who is imagining what it would be like to live in the same health state. Another argument for using the preferences of the general population is that this population finances health care and should set priorities. Organisations using cost-utility studies for evaluation, such as NICE, have stated that utilities from the general population are to be used. The EQ-5D fulfils this requirement since the tariff was established using general population preferences.

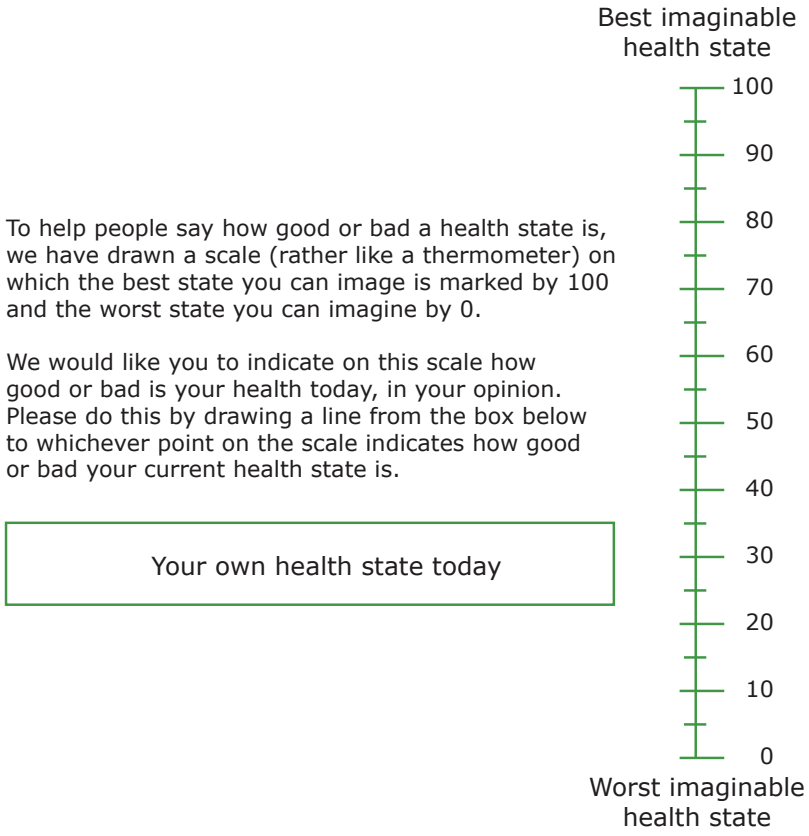
The EQ-5D appears currently to be the most frequently-used instrument because, as a simple generic questionnaire, it can be used in virtually any study, be it a cross-sectional survey, a patient registry, or a clinical trial. In addition, it has been translated into a large number of languages. The EQ-5D is a two-part measure that provides both a compact descriptive profile and a single index value. For utility assessment, the descriptive portion is used. This includes five dimensions of health (see Figure 3.14) at three levels of severity, coded as 1 (no problems), 2 (some problems), and 3 (severe problems). A refined version with five levels is being developed (Herdman et al, 2001; Devlin and Krabbe, 2013) and research is underway to produce tariffs for this new instrument.

For the three-level version, an individual's combination of attributes and levels are used to define one of 243 theoretically possible health states. For example, state 11111 indicates no problem in any of the domains, and is equivalent to full health. The EQ-5D also includes a visual analogue scale in the form of a health thermometer (see Figure 3.15).

Figure 3.14. The five dimensions of the EQ-5D

Dimension	Level
Mobility	No problems walking about
	Some problems walking about
	Confined to bed
Self-care	No problems with self-care
	Some problems washing or dressing self
	Unable to wash or dress self
Usual activities	No problems with performing usual activities (e.g. work, study, housework, family or leisure activities)
	Some problems with performing usual activities
	Unable to perform usual activities
Pain/discomfort	No pain or discomfort
	Moderate pain or discomfort
	Severe pain or discomfort
Anxiety/depression	Not anxious or depressed
	Moderately anxious or depressed
	Severely anxious or depressed

Figure 3.15. The visual analogue scale in the EQ-5D: the EQ-VAS



When it is possible to elicit preference weights directly from patient groups in an interview, two main techniques may be used.

1. The standard gamble (SG), diagrammed in Figure 3.16, is a classic technique that has been extensively used in decision analysis, e.g. to assess the closely related issue of risk aversion. It uses the axioms of expected utility theory to measure the utility that an individual attaches to any given health state.

2. The time trade off (TTO) method, shown in Figure 3.17, was developed in the early 1970s specifically for use in health care.

Box 3.11. Standard gamble

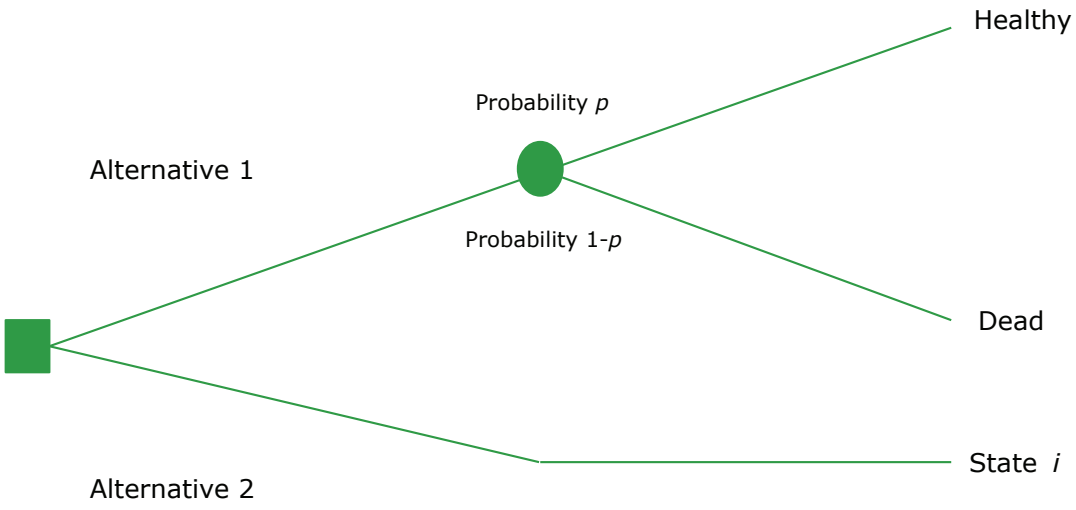
In a typical standard gamble scenario, an individual is offered two alternatives.

Alternative 1 has two possible outcomes: either return to full health for the remaining years of life expectancy with a probability of p, or experience immediate death with a probability of 1-p.

Alternative 2 has one certain outcome of a chronic health state i for the remaining years of life expectancy.

The individual is then allowed to vary the probability, p, until they are indifferent to the two alternatives. If full health and death are automatically assigned utilities of one and zero, respectively, then the utility for state i is given by p. At the point of indifference between the two alternatives, $p(1) + (1-p)(0)$ = utility of health state i. The utility of health state i therefore = p.

Figure 3.16. Standard gamble



Box 3.12. Time trade-off

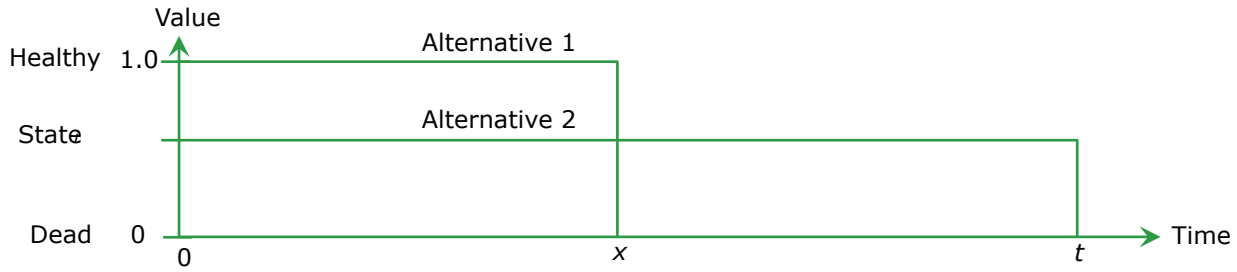
As in the standard gamble, the time trade-off method offers two alternatives.

Alternative 1 is full health for time x followed by death.

Alternative 2 is to remain in intermediate health state i for time t ($t > x$), followed by death.

The individual is then allowed to vary time x until they are indifferent to the two alternatives. If full health and death are assigned utilities of one and zero, respectively, then the utility for health state i is given by x/t. At the point of indifference between the two alternatives, $x(1) = t$? utility of health state i. Therefore, the utility of health state $i = x/t$.

Figure 3.17. Time trade-off

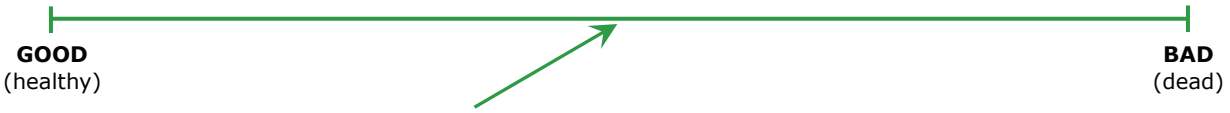


Visual analogue scales (VAS) (see Figure 3.13) sometimes are the simplest method for eliciting health state utilities. Individuals are asked to indicate where on a line between the best and worst imaginable health states they would rate a pre-defined intermediate health state. The EuroQol-VAS thermometer (Figure 3.15, above) is an example. This method is not preferred, however, because scores on a VAS are generally much lower than those obtained using SG or TTO. This is because respondents use the full scale from 0 to 1 without any other consideration of risk of loss.

Box 3.13. Visual analogue scale

With visual analogue scales, individuals are asked to indicate where on the line between the best and worst imaginable state they would rate a certain health state, either their own or one that is described to them. The health state valuation is then derived by measuring the distance between healthy (generally assigned 1) or dead (generally assigned 0) and the indicated health state on the line. For example, on a 10 cm line with death at 0 and full health at 10cm, a health state indicated as being located 8cm along the line would receive a score of 0.8 (8/10).

Figure 3.18. Visual analogue scale (VAS)



The basic techniques of a cost-utility analysis are shown with a hypothetical example in Box 3.14.

Box 3.14. Theoretical example of cost utility analysis

Treatment A improves survival by 1 year with a quality of life (utility) of 0.7 at a cost of €1,400.

Treatment B improves survival by 1.2 years, but with a lower utility of 0.6 at a cost of €2,160.

The average cost per utility of A is €2,000 per QALY: €1,400 / (0.7 x 1)

The average cost per utility of B is €3,000 per QALY: €2,160 / (0.6 x 1.2)

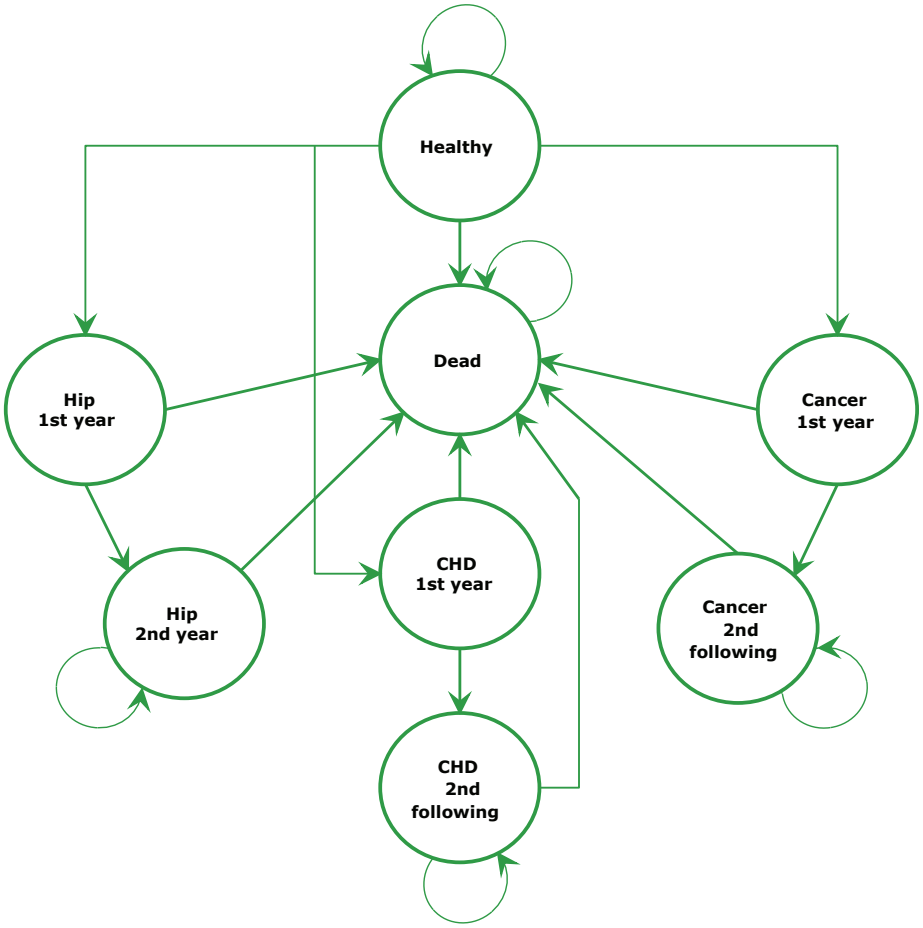
The incremental cost-utility of B over A is €38,000 per QALY: (2,160-1,400) / (0.72-0.70)

Study Example 7. Cost-utility analysis: incorporating multiple events—osteoporosis

Whenever research and clinical practice produce new knowledge about a disease and new data become available, the cost-effectiveness of treatment strategies is likely to change and cost-effectiveness models must be updated. Osteoporosis is an example where models evolved from working with a surrogate endpoint—bone mineral density (BMD)—to estimating events avoided (fractures), and then to using QALYs to incorporate mortality and related events. The study presented here (Zethraeus, Johannesson and Jönsson, 1999) shows how more than one type of event can be incorporated into a single model and lifetime QALYs can be estimated. The example also illustrates the complexity of such models and, even more so, the large amount of data that is required.

Osteoporosis is characterised by low bone mass, which implies that the risk of fractures is greater. Common osteoporosis-related fractures are those of the spine, wrist and hip, which can lead to reductions in the individual's quality of life both at the moment of the fracture and in the long term. Hip fractures, particularly in the elderly and frail, also carry a risk of mortality. Costs that can be attributed to osteoporosis are resources used for prevention, the treatment of fractures, and rehabilitation.

Figure 3.19. Model structure for hormone replacement therapy



Early economic evaluations of preventive treatments were based on changes in BMD, which was then linked with an epidemiological risk function for hip fractures at a given age and level of BMD. Since the early 1990s, fracture studies, usually of the spine and/or wrist, are required to receive licensing approval for a treatment. The cost of events avoided (fractures) could be calculated directly. The reason to prevent fractures, however, is to avoid the morbidity and mortality that is associated with them in the long term. Therefore, the cost per event avoided neither reflects the true outcome nor is helpful in decisions on budget allocation. In addition, current treatments for osteoporosis can affect multiple endpoints, such as different types of fractures, or have extra-skeletal consequences. An

example of the latter is hormone replacement therapy (HRT), thought not only to reduce the risk of fractures, but also to decrease the risk of coronary heart disease (CHD) and, conversely, increase the risk of breast cancer.

To estimate the cost-effectiveness of such an intervention requires a model that includes the risks of all relevant events and their impact on morbidity and mortality. The model in Figure 3.19, above, was built to estimate the cost-effectiveness of HRT for preventing fractures in post-menopausal women. The overall structure and the Markov states are shown.

Table 3.9. Costs used in the model (1998 USD)

Age	Cost type	AMI (recog.)	AMI (unrecog.)	Angina pectoris	Coronary insufficiency	Hip fracture	Breast cancer	Cost LYG
50	Direct	6,250	437	6,250	10,625	9,875	8,375	-4,625
	Indirect	11,250	3,437	11,250	11,250	10,000	10,375	—
65	Direct	6,250	437	6,250	10,625	10,750	8,375	19,875
	Indirect	0	0	0	0	0	0	—
75	Direct	6,250	437	6,250	10,625	18,875	8,375	19,875
	Indirect	0	0	0	0	0	0	—
85	Direct	6,250	437	6,250	10,625	26,375	8,375	19,875
	Indirect	0	0	0	0	0	0	—
			Second and following years					
50	Direct	875	437	875	875	5,125	150	—
50	Indirect	6,875	3,437	6,875	6,875	7,000	175	—
65+	Indirect	0	0	0	0	0	0	—

Key: AMI = acute myocardial infarction (recognised and unrecognised); LYG = life-year gained
Source: Zethraeus, Johannesson and Jönsson (1999)

Table 3.10. Utility weights used in the model

Age	Cardio-vascular disease	Hip fracture	Breast cancer	Population
	First year			
50–64	0.80	0.70	0.80	0.90
65–74	0.69	0.59	0.69	0.79
75+	0.53	0.43	0.53	0.63
	Second and following years			
50–64	0.80	0.80	0.80	0.90
65–74	0.69	0.69	0.69	0.79
75+	0.53	0.53	0.53	0.63

* Includes AMI (recognised and unrecognised), angina and coronary insufficiency

Each state is associated with age-dependent mortality rates, costs and utility weights. The cycle duration for the model is one year. The disease states are divided into "first year" and "second and

following years" after a disease event because mortality rates, costs, and utility weights differ between these time periods. Costs and utility weights for the different states are shown in Tables 3.9 and 3.10, respectively.

The basic model structure assumes a cohort of healthy individuals, i.e. free of CHD and breast cancer and with no previous fracture. After each cycle, the cohort is reallocated to the different health states according to the transition probabilities. In the first cycle, the cohort is exposed to the risks of CHD, breast cancer, hip fractures and death from other causes. When a patient experiences an event, only transitions to "second and following years" (post-event state) or death are relevant. Patients in the second and following years' states remain in these states until they die. The cohort is followed to age 110 years. The cost-effectiveness formula used in the computer model can be expressed as follows.

$$\frac{\Delta C}{\Delta E} = \frac{C1 - C0}{E1 - E0} = \frac{\Delta INT + \Delta MORB + \Delta MORT}{\Delta QLE} = \frac{\Delta INT + \Delta MORB + \Delta MORT}{\Delta LE + \Delta LEQ}$$

where

ΔINT = intervention costs, direct and indirect

ΔMORB = changes in morbidity costs (direct, indirect) due to the intervention

ΔMORT = changes in mortality costs (direct, indirect) due to the intervention

ΔLE = changes in life expectancy due to the intervention

Δ LEQ = changes in quality of life measured in years due to the intervention

Δ QLE = ΔLE + ΔLEQ

The model allows for the inclusion of costs in added life years, and results are expressed either as costs-per-life-year gained (LYG) or costs-per-QALY gained. As the model incorporates consequences for different diseases, a composite outcome measure (e.g. QALY) is needed because it has the capacity to incorporate the intervention's effectiveness for different risks from different diseases.

Costs in all health states include direct and indirect costs. Annual intervention costs include the cost of drugs, cost of services in hospitals and primary health care, cost of travelling and the indirect cost of production forgone due to the treatment. In addition, the intervention has "initial costs" such as those for screening or diagnosis. Changes in morbidity costs are costs saved because of reduced morbidity from CHD and hip fractures adjusted for costs added, because of increased morbidity from breast cancer. Changes in mortality costs are equal to changes in total consumption minus production that are due to changes in mortality because of the intervention.

An example of using this model is investigating the cost-effectiveness of HRT given to asymptomatic women for ten years in Sweden, shown in Table 3.11. Depending on menopausal status and age (50, 60 and 70), six independent treatment groups were identified. The annual average intervention cost was estimated at SEK 2,000.

Table 3.11. Cost-utility of ten-year intervention with HRT in asymptomatic women: costs per life year and QALY gained (QALY in parentheses), in SEK '000s

Risk change	Oestrogen (women with hysterectomy)			Oestrogen +progestogen (intact uterus)		
	Age			Age		
	50	60	70	50	60	70
Hip -40%, CHD -20%	400 (310)	240 (230)	170 (190)	580 (450)	300 (300)	200 (230)
Hip -40%, CHD -50%	160 (140)	170 (190)(200)	160 (200)	230 (220)	200 (220)	180
Hip -50%, CHD -20%	360 (280)	210 (200)	150 (170)	540 (410)	280 (260)	180 (200)
Hip, -50%, CHD -50%	150 (120)	160 (170)	150 (180)	220 (190)	190 (200)	170) (200)
Hip -40%, CHD -20%, Cancer+35%	D (640)	270 (240)	170 (190)	D (1,060)	370 (320)	210 (230)
Hip -50%, CHD -20%, Cancer+35%	190 (130)	180 (180)	160 (200)	320 (230)	210 (220)	180 (220)
Hip -50%, CHD -20%, Cancer+35%	D (500)	240 (200)	150 (160)	D (860) (860)	330 (280)	180 (200)
Hip -50%, CHD -50%, Cancer+35%	170 (120)	170 (170)	150 (180)	300 (210)	200 (200)	170 (200)(

Key: D means that HRT is dominated by the "no intervention" alternative
Source: Zethraeus, Johannesson and Jönsson (1999)

Study Example 8. Cost-utility analysis in chronic progressive diseases—rheumatoid arthritis

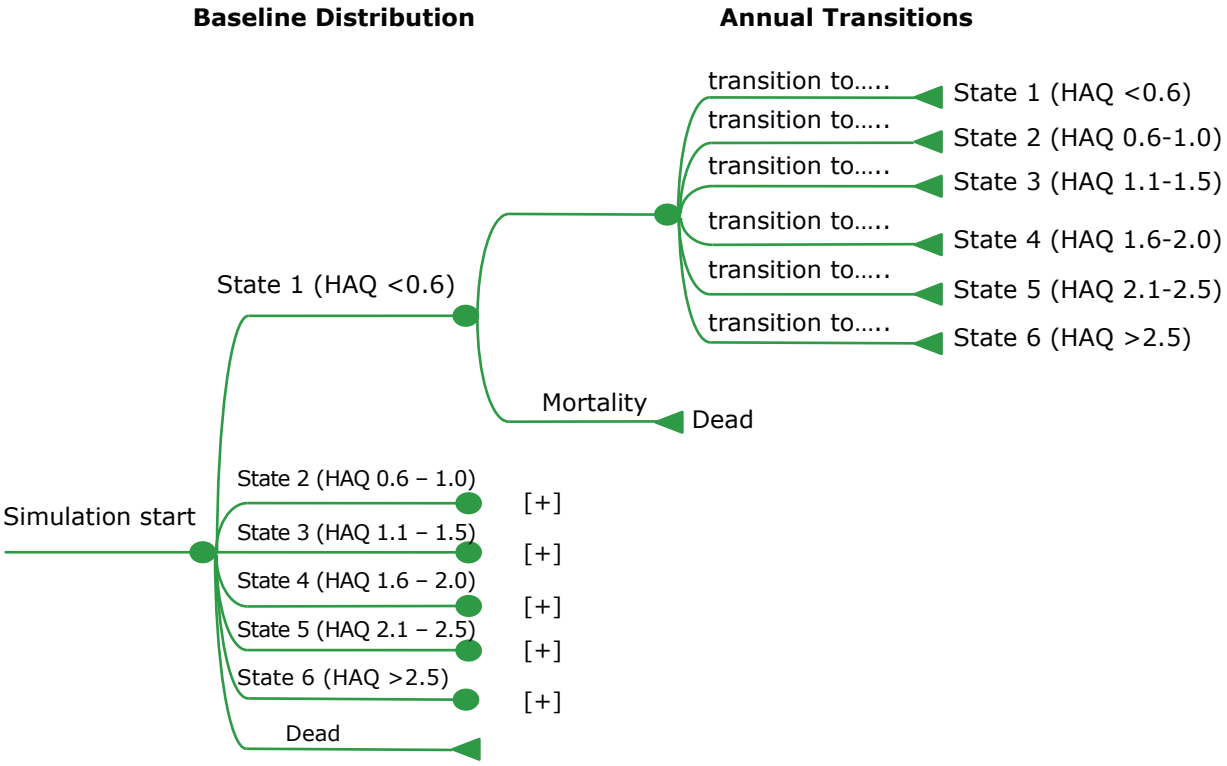
Studies in RA illustrate a number of issues the health economist faces: modelling a progressive disease with multiple symptoms; extrapolating data from short term clinical trials to the long term; adjusting models to new data, as well as to a changing market where multiple similar products are available; and working with patient registries from routine clinical practice, both for new studies and for verifying results based on clinical trials versus real life data. These issues are illustrated here with a general model in RA and four adaptations of the original model.

In fields such as osteoporosis or cardiovascular disease, the goal of treatment is to control a risk factor to avoid a future event, such as control of BMD to avoid osteoporotic fractures or of blood pressure and cholesterol levels to avoid cardiac events. In those cases, economic evaluation can be based on distinct endpoints. In chronic progressive diseases, such as RA or MS, however, the goal of treatment is to alleviate diverse symptoms and slow progression of the disease. No obvious clinical endpoint exists that can serve as an outcome for economic evaluation. The burden of these diseases comes from the fact that onset is relatively early in life and patients thus live with the disease for a long time, with progressing disability and detrimental effects on quality of life. Because economic evaluation must capture both symptoms and progression, the QALY appears to be the most appropriate and comprehensive measure of effectiveness.

Typically, clinical trials are too short to estimate the benefit of slowing progression to severe health states that create disability or reducing mortality associated with a slow degenerative disease. Trials for RA can measure the effect of treatment for reducing inflammation, which will be apparent within a

few weeks and well within the time frame of a clinical trial. The effects measured in the trial must then be extrapolated into the future. This requires a model that describes the disease progress in terms that are relevant for economic evaluation, for instance, with a measure that can be related to both quality of life and costs over time. An additional requirement is that the measure must be used widely in both epidemiological cohorts and clinical trials.

Figure 3.20. Structure of the RA models



Source: Kobelt et al (2002)

One solution is a Markov model. A number of such models have been published for RA as more epidemiological data have become available. An early study demonstrated how such a general disease model could be developed for RA and used to calculate the cost-effectiveness of treatments that affect symptoms and disease progression (Kobelt et al, 1999). The initial five-year model was updated when 15-year epidemiological follow-up data became available (Kobelt et al, 2002) and was further refined as information on the disease became more comprehensive (Kobelt et al, 2005a). These models were then used to estimate the cost-effectiveness of biologic treatments compared to previous standard care in RA or to other biologics.

Four examples are included here: the first, infliximab illustrates the issues with extrapolation when no long term data are available and also how to model the placebo effect (Kobelt et al, 2003); the second, etanercept, illustrates how multiple disease measures with different effects on costs and utility can be combined to refine the analysis (Kobelt et al (2005b); the third illustrates how registry data can be used (Lindgren, Geborek and Kobelt, 2009). In the fourth, the results of the early cost-effectiveness analyses of infliximab were verified against real world data from registries (Lekander et al, 2013).

The original model was based on a Swedish cohort study where 183 patients with early RA were followed for up to 15 years. Although both functional and radiological measures were available to define Markov states, only functional data were used, measured with the Health Assessment Questionnaire (HAQ). The HAQ is used in all clinical trials in RA, as well as in cohort studies, and is a patient-centric measure rather than a "hard" clinical endpoint, such as X-ray based radiological scores. Also, major changes in joint erosion, particularly in larger joints, will seldom be quantifiable in short-term clinical trials and so cannot provide data for economic evaluation of new treatments.

Finally, the HAQ has been found to correlate well with resource consumption and quality of life (utility) while radiological scores do not.

The structure of the model is shown in Figure 3.20. HAQ scores were initially grouped into six states and a state for death (normal mortality) was added; in subsequent models, the two most severe states were regrouped because patient numbers in these states in available datasets were small. Cycles in the model are one year with disease progression based on the epidemiological study and modelled as annual transitions between the states, conditional upon time (i.e. annual cycles elapsed) and patient characteristics such as age, sex, and time since disease onset, using a probit model.

Table 3.12 illustrates the progression of the disease over ten years in the Swedish cohort study. Over time, more patients can be found in the more severe states 3 to 6, and fewer in the milder states 1 and 2 (excluding those patients who died).

Table 3.12. Cohort distribution over time (percentage of patients)

Disease State	Percentage of patients									
	Year									
	1	2	3	4	5	6	7	8	9	10
1	34	36	36	35	26	29	34	32	30	28
2	38	31	31	28	40	33	24	25	25	28
3	14	25	25	23	26	27	24	27	23	23
4	9	4	5	10	5	8	16	12	17	13
5	4	3	3	3	2	2	2	4	4	4
6	1	0	1	1	0	0	0	1	0	3

Source: Kobelt et al, 2002

Costs and utilities differ by state, but are constant for all patients within the same state, irrespective of age, gender or other factors. Direct medical costs as well as sick leave and early retirement were available directly from the cohort study and incorporated in the first models. However, observational studies subsequently showed the importance of direct costs borne by patients and their families, such as special equipment, home adaptation and informal care. Consequently, cost data were incorporated into the models from a different source, an observational study in Sweden, which also provided utility data.

A further refinement concerned the impact of inflammation and pain (disease activity) on costs and utilities. Disease activity has taken on a key role in patient management in recent years and is currently measured in both clinical trials and clinical practice, in particular to define treatment effect. It should therefore be included in treatment models. Analysis of a dataset where costs, utilities and HAQ score, as well as disease activity scores (DAS) were measured, showed that DAS had an additional effect on utilities when controlling for HAQ score. This means that patients within the same HAQ state in the model with high and low DAS had different utilities, making it necessary to subdivide each HAQ state. DAS was also a strong predictor of sick leave (HAQ was not), but had no correlation with other costs when controlling for HAQ. Costs in the model were modified to account for this. Table 3.13 illustrates mean annual costs and utilities by Markov state.

Table 3.13. Mean costs per cycle and utilities for different Markov states (2004 EUR)

	Mean annual costs					Mean utilities		
State	Medical costs	Patient costs	Work absence	Early retire	Total	Utility (high DAS)	Utility (low DAS)	Mean utility
1	773	276	1,311	1,882	4,242	0.709	0.780	0.768
2	1,590	421	2,629	4,310	8,950	0.568	0.704	0.645
3	2,456	760	2,496	6,467	12,179	0.441	0.676	0.539
4	3,496	1,388	1,142	7,468	13,949	0.446	0.562	0.488
5	8,890	1,333	470	7,902	18,595	0.213	0.408	0.239

Source: Kobelt et al (2005a)

Study Example 8.1

In the first example of cost-effectiveness for RA, the economic evaluation was based on a one-year study comparing infliximab to placebo, both with background methotrexate, with no data available beyond one year. Thus, the study only incorporated treatment and outcome for the duration of the trial and assumed that the outcome achieved during the trial would be either maintained for some time (Model A below) or lost after the trial (Model B below). Since stopping treatment after one year clearly does not represent routine clinical practice, the model thus represents the trial only and limits the number of other assumptions that would have been necessary if treatment had been assumed to continue. These relate to the continued effect of treatment on inflammation and on the development of HAQ, as well as to discontinuation and adverse events. These assumptions have subsequently been compared to real life data (see Study Example 8.4).

Two different methods were used to estimate cost-effectiveness.

1. HAQ scores from the trial were used in the first cycle of the model for both groups and the disease model described above was used to extrapolate to ten years (Model A). This scenario assumes that the treatment effect is maintained for some time and incorporates the placebo effect in the same way as the treatment effect.

2. The difference between the treatment and placebo groups is used to calculate odds ratios of worsening or improving; these are then applied to the cohort study (Model B). This scenario allows the elimination of the placebo effect (Model B-1), which increases costs in the placebo arm, and the exploration of different rates of loss of effect at treatment discontinuation (Model B-2), which increases costs in the treatment arm.

Both models only use HAQ to estimate utilities.

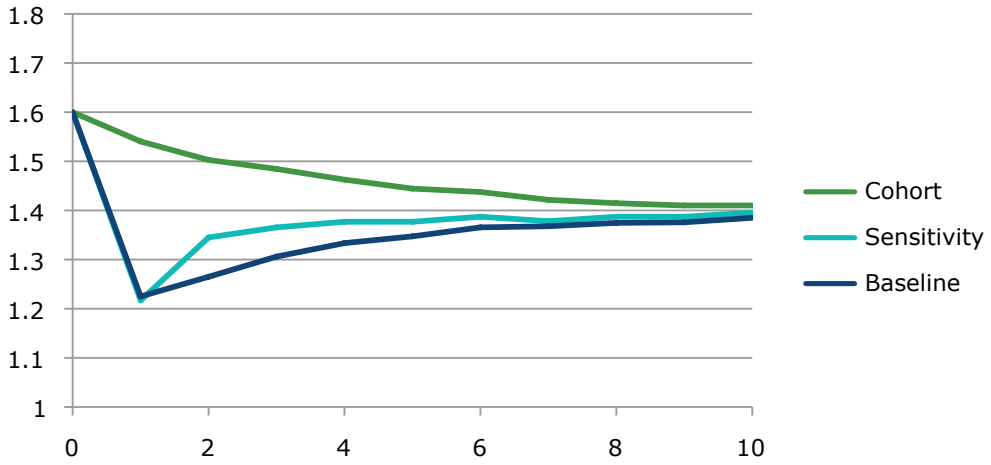
Table 3.14 summarises the results of the analysis for both models. Figure 3.17 illustrates the development of functional capacity in the model under different assumptions regarding the maintenance of the trial effect. Simulations are run for ten years and discounted at 3%.

Table 3.14. Cost effectiveness scenarios with infliximab (SEK, discounted 3%)

Scenario	Infliximab	Placebo	Difference
Model A (one year treatment)			
Direct cost	257,826	191,857	65,969
Total cost	1,129,507	1,121,476	8,031
Utility	4.632	4.384	0.248
Cost/QALY, direct costs			266,000
Cost/QALY, all costs			32,000
Model B-1 (one year treatment)			
Direct cost	266 757	212 391	54,366
Total cost	1,187,780	1,250,406	-62,626
Utility	4.648	4.417	0.231
Cost/QALY, direct costs			235,000
Cost/QALY, all costs			(cost-saving)
Model B-2 (one year including loss of effect at treatment discontinuation)			
Direct cost	270,774	212,391	58,383
Total cost	1,219,365	1,250,406	-31,041
Utility	4.596	4.417	0.179
Cost/QALY, direct costs			325 000
Cost/QALY, all costs			(cost-saving)

Source: Kobelt et al (2003)

Figure 3.21. Average development of HAQ scores in the model, adjusted for different effectiveness after a one-year trial



Note: Lower HAQ values indicate less disability

Key: Cohort: transitions from the epidemiological study. Base case: the treatment effect of infliximab (Remicade®) is maintained when treatment is stopped and patients follow the transitions of the epidemiological cohort. Sensitivity analysis: some of the treatment effect is lost within the year following treatment cessation, after which patients follow the transitions of the epidemiological cohort.

Source: Kobelt et al (2003)

Study Example 8.2

The second example for RA is based on a two-year clinical study comparing etanercept monotherapy, the combination of etanercept with methotrexate, and methotrexate alone in patients with active disease. The modelling study (Kobelt et al, 2005a) was performed using data from the open extension phases from previous etanercept clinical studies, enabling extrapolation beyond the trial. Indeed, these follow-up studies indicated that the treatment effect was maintained while remaining on treatment, but lost over a period of some months following discontinuation of treatment. The economic analysis was performed for both: treatment in the trial period only (as in the previous example) and with treatment extrapolated for the full duration of the simulations. In the former case, the treatment effect was lost entirely during the following cycle; in the latter case, only patients discontinuing treatment lost the effect while patients on treatment remained almost stable. In addition to HAQ data, this model included the effect of disease activity and radiographic progression on costs and utilities. Simulations were again run for ten years in order to limit the uncertainty, and costs and effects discounted were at 3%.

Table 3.15 shows the results of both analyses. The estimates in the first case illustrate the concept of "dominance": Although etanercept monotherapy had a better effect than methotrexate (but also higher costs, thereby yielding an ICER of €123,850 per QALY gained), the combination of etanercept and methotrexate was both less expensive and more effective than etanercept alone. Etanercept monotherapy is thus dominated by the combination therapy and from an economic point of view should not be recommended. This does not mean, however, that from a clinical point of view, monotherapy should not be considered for certain patients. Consequently, the cost-effectiveness of the combination is established by comparing to methotrexate alone, excluding etanercept monotherapy, as shown in the second case.

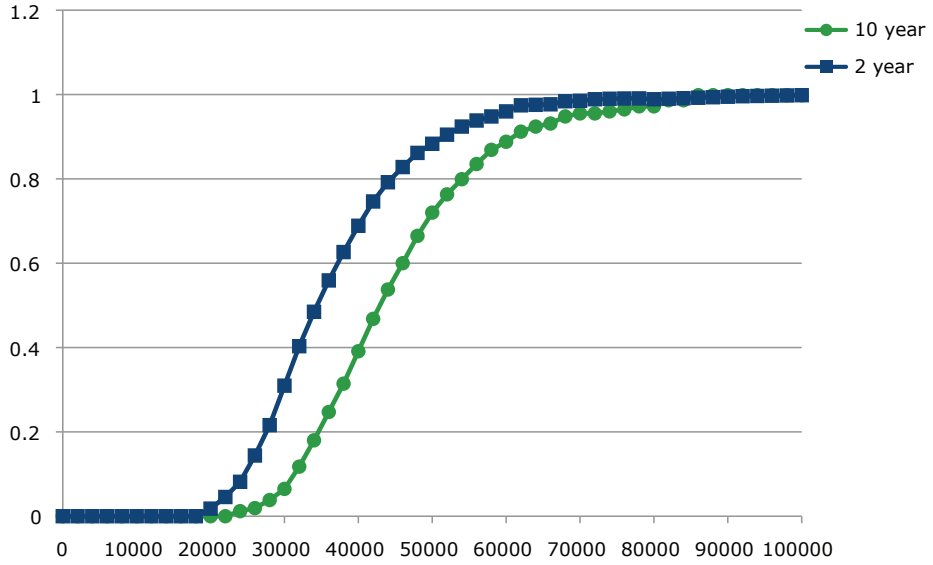
Table 3.15. Cost per QALY gained with etanercept in Sweden (2004 EUR, discounted 3%)

	Costs	Difference to methotrexate	Effects (QALYs)	Difference to methotrexate	Cost/QALY vs methotrexate
Treatment for 2 years, extrapolation to 10 years					
Methotrexate	162,695	—	3.08	—	—
Etanercept	181,271	18,577	3.23	0.15	Dominated by the combination
Etanercept+ methotrexate	176,915	14,221	3.46	0.38	37,331
Treatment for 10 years					
Methotrexate	149,943	—	3.43	—	—
Etanercept+ methotrexate	192,091	42,148	4.34	0.91	46,494

Source: Kobelt et al (2005b)

This example also illustrates the use of acceptability curves discussed earlier (see Figure 3.18, above.). The curves were generated by Monte Carlo simulation, using 1,000 individually-drawn bootstrap estimates from the entire distribution of costs and utilities in the different Markov states, rather than the mean values presented in Table 3.13 and used in the cohort simulations in Tables 3.14 and 3.15. The resulting curves thus allow us to get an indication of the uncertainty in the estimates: If the willingness to pay for a QALY gained is €50,000, there is an 88% probability that a two-year treatment with etanercept plus methotrexate is acceptable in this type of patient; for a ten-year treatment, the probability is 71%. If the threshold value were €82,000, the probability would be 99%.

Figure 3.22. Uncertainty in ICERs



Source: Kobelt et al (2005a)

Study Example 8.3

Both these early cost-effectiveness studies in RA compared a biologic treatment with the most effective non-biologic therapy available, methotrexate. Despite a large difference in the drug cost (in Study Example 8.2, the annual cost was €78 for methotrexate and €16,000 for etanercept), the biologic treatments were judged acceptable. In recent years, a large number of other biologic drugs have been introduced for RA, changing the focus of economic evaluations. The economic question no longer is whether to use a biologic or non-biologic treatment, but rather what treatment strategies should be adopted for patients who qualify for these agents, i.e. which biologic should be used first, which is suitable for second-line therapy depending on whether the patient cannot tolerate or does not benefit from the first-line therapy, and which is suitable for third-line therapy, etc.

This requires adjusting the economic models by adding yet another feature. In addition to epidemiological data on progression, and costs and utilities related to function and disease activity, data on discontinuation rates, and switching to other treatments and the effectiveness of different treatments are needed. Patient registries can provide this information from clinical practice, as illustrated by the analysis of rituximab described in this example (Lindgren, Geborek and Kobelt, 2009).

The clinical trial compared rituximab to placebo, both with background methotrexate, in patients who had had an inadequate response to one of the three biologics available at that time. All three treatments were in the same class of drugs (TNF-inhibitors) and, following the failure of the biologic, a treatment with a different mechanism of action was expected to be more effective than using a second, similar treatment. In the economic evaluation, rituximab was therefore compared to the mix of treatments used in second-line therapy.

The biologics registry from southern Sweden provided patient level data on first-, second- and third-line use of TNF-inhibitors (1,903, 633 and 170 patients, respectively) for this analysis: time on and between treatments, discontinuation, adverse events, development of function (HAQ), disease activity (DAS) and utility scores (EQ-5D, 6860 observations). Based on these data, the model was redeveloped as a discrete event simulation (DES) model, using the time to events defined as "start treatment", "stop treatment" and "die". Patients would start with a first-line TNF-inhibitor; discontinue; start the second immediately, after some time, or not at all; discontinue; and restart with the third immediately, after some time, or not at all.

Patients could die at any moment, according to normal and RA-specific mortality. Each of the health states in the model was associated with the changes in HAQ and DAS status over time, which was estimated from the registry and then converted to costs and utilities. To estimate the cost-effectiveness of rituximab, the second TNF-inhibitor in the model was replaced with rituximab using the data from the clinical trial. The analysis thus compared a strategy where rituximab was used in second-line therapy to one with a sequence of TNF-inhibitors only. This poses the problem of comparing clinical practice and clinical trial data, which is unavoidable in this situation and is best handled with extensive sensitivity analysis.

Simulations were performed for a population matching the trial population over its expected lifetime, adopting a societal perspective with costs and utilities discounted at 3%. For the deterministic analysis, a female patient with the mean age, HAQ, DAS and disease duration from the trial was used. The base case predicted a mean survival of 24 years, and results favoured rituximab, with a cost saving of €2,500 and an incremental effect of 0.2 QALYs gained. Cost differences are mostly explained by the lower price of rituximab and effect differences by the fact that patients spend more time on treatment because a fourth option is available with rituximab.

Probabilistic sensitivity analysis is presented differently from the previous example: 1,000 second-order simulations of the cost-effectiveness ratios are presented as a scatter plot with the ellipse indicating the 95% confidence interval (i.e. covering 95% of the simulations) and the dotted line a theoretical willingness to pay for a QALY of SEK 500,000 (€53,000). As can be seen, all but one of the simulations is acceptable if willingness to pay is SEK 500,000.

Study Example 8.4

Analyses performed at introduction of a new treatment are by necessity based on efficacy from randomized clinical trials, rather than on effectiveness in clinical practice. This means that a number of assumptions must be made, which produces substantial uncertainty in the results. The study in Study Example 8.1 is a typical illustration of this situation, with the main necessary assumptions relating to the effect of treatment after the trial, the discontinuation rate, the potential loss of effect at discontinuation and most importantly, the comparator. One therefore should routinely verify the results once real world data become available and a number of authorities now request such studies, e.g. when drugs are re-evaluated for reimbursement. Sweden is among those countries.

The model used for the analysis of the ATTRACT trial was thus populated with data from the Swedish RA registry, and three patient populations were compared:

- The original ATTRACT patients (N=340)
- Patients treated with infliximab in the registry (N=637)
- A subset of the registry patients that would have fulfilled the enrolment criteria for ATTRACT (matched cohort, N=306).

The key differences in the assumptions based on the clinical trial and the data from the registry related to:

1. Comparator: placebo in the early analysis, natural disease history in the verification
2. Discontinuation: in the trial, all patients discontinue at the end of the trial; discontinuation in the registry is as observed over ten years
3. Loss of effect on discontinuation: maintenance of effect achieved versus loss of effect within one year

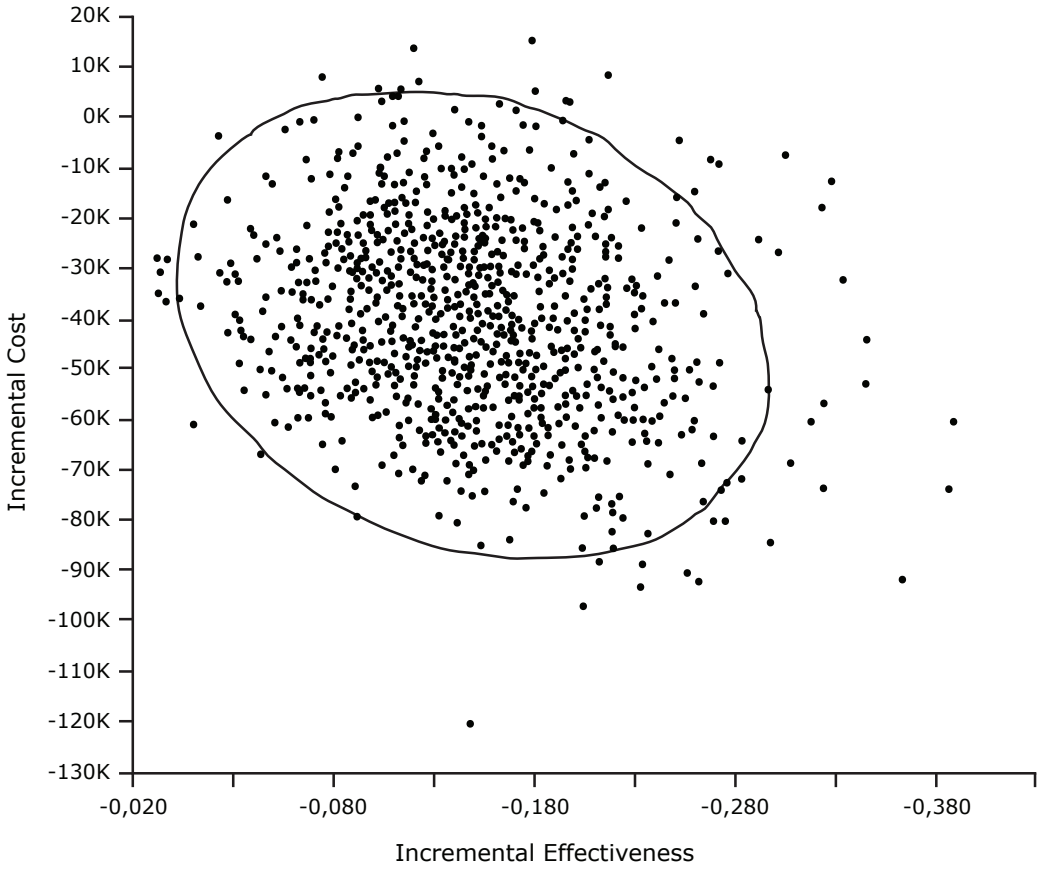
4. Additional effect of disease activity: not included in the early model, added to the model (see Study Example 8.2)

The analysis showed that the results were identical when the ATTRACT population and the matched real-world cohort were compared using the same assumptions. Thus, the model predicted correctly for patients such as those included in the trial. When the effect of disease activity was included, the cost per QALY gained decreased to half; replacing placebo as comparator with natural history resulted in cost-savings.

The greatest change in the results came, as expected, from discontinuation. When treatment as observed in the registry was included, with a loss of effect after discontinuation, the QALYs gained increased, but so did costs, resulting in a higher cost per QALY. Surprisingly, differences were legible when all patients in the registry, not only the matched cohort, were included. This indicates that the type of patient treated with infliximab in the early years corresponded to the trial.

In conclusion, the study showed that early models can indeed predict cost-effectiveness in clinical practice, with the caveat that major assumptions must be carefully evaluated as they can have a large effect on the results.

Figure 3.23. Probabilistic sensitivity analysis



Source: Lindgren, Geborek and Kobelt (2009)

Study Example 9. Cost-utility analysis in chronic progressive diseases—multiple sclerosis

This example further illustrates the issues for economic evaluation of treatment for chronic progressive diseases. Compared to RA, the treatment effect for MS is more difficult to measure. The onset of MS occurs at a young age, generally around 30 to 35, and disease progression is very slow, spanning several decades. However, exacerbations occur at an average of 0.5 to 1 per year and, therefore, constitute an outcome that is measurable in clinical trials. The series of four studies described shows again how economic models need adapting as knowledge and the market environment change. It also reinforces the fact that changes to models are driven principally by the type and quality of data available including effectiveness data from clinical practice registries.

Compared to RA, MS is a more difficult disease to model. It shares all of the issues of RA, but symptoms are more diverse. In addition to progression of disability, distinct events (exacerbations) occur, particularly in the early phases of the disease, creating extreme functional disability for a period of time. This has produced two main definitions of the disease, relapsing-remitting MS (RRMS) and secondary progressive MS (SPMS), although transition from RRMS to SPMS is not well defined.

Since the mid-1990s, new medicines to treat MS have been licensed for use in patients with RRMS because they reduce the number and severity of relapses, although their effect on progression is less clear. However, the primary objective of treatment for MS is to avoid or slow progression to the severe stage of the disease. Evidence suggests a link between the number and frequency of relapses and disease progression, but no definitive dataset has been produced that has enabled the calculation of a risk function for this link. Economic evaluations, therefore, need to consider the entire course of the disease. This poses the question as to whether two clearly distinct phases of a disease, as well as the link between the two, should be modelled or whether the disease should be considered as a continuum.

In the study series presented here, two distinct phases were initially modelled because one trial had shown a significant effect in both RRMS and SPMS. In view of the large overlap of RRMS and SPMS in the mid-range of the functional scale used to assess disease progression (Extended Disability Status Scale, or EDSS), the phases were then combined. Other studies have maintained the separation, as theoretically only patients with RRMS should be treated. However, relapses continue during the progressive phase, albeit with a lower frequency, and registry data have shown that in practice treatment continues even after the conversion to SPMS.

Table 3.16. Cost and utilities by levels of disability in Sweden in EUR (2005)

EDSS score	Direct costs	Informal care costs	Indirect costs	Utility
0-1	1,813	406	4889	0.825
2	8,457	1,065	11,638	0.696
3	6,142	1,747	18,757	0.646
4	12,063	1,627	12,774	0.610
5	15,458	3,406	21,100	0.583
6	13,546	4,297	20,422	0.572
6.5	21,515	6,322	25,826	0.462
7	37,553	7,113	27,247	0.373
8-9	77,574	12,061	33,144	0.047

* Excluding DMDs **Applied to patients less than age 65

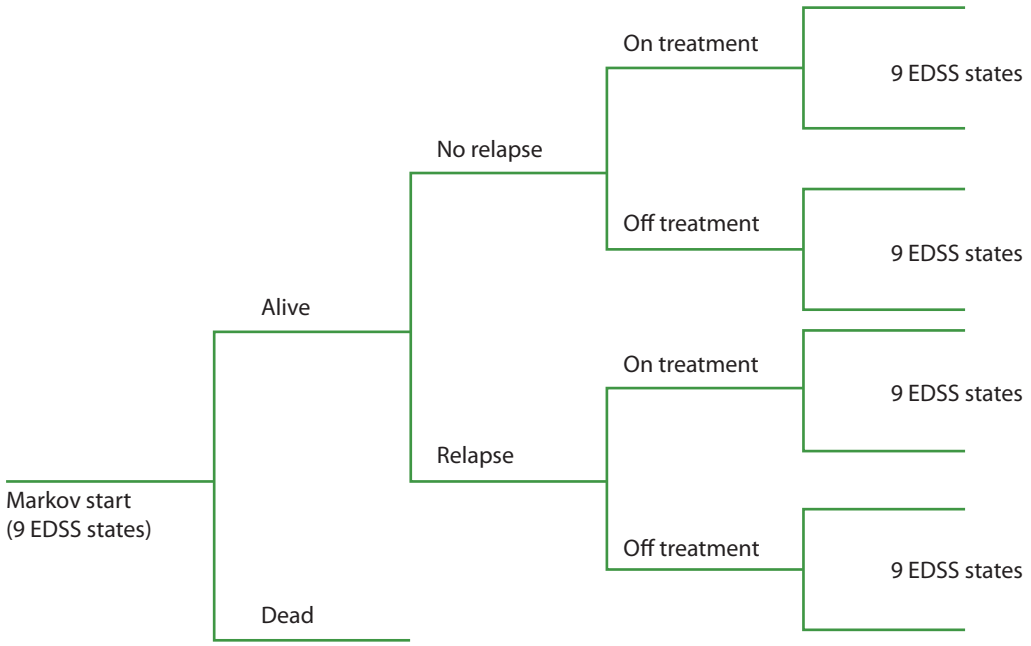
Source: Kobelt et al (2008a)

The three early versions of the model used data on resource utilisation and utilities that were collected in the first observational study series, while the fourth version includes data from the more recent series presented (see Study Example 1, chapter 2). The first version (Kobelt et al, 2000) was based entirely on clinical data; the second version (Kobelt et al, 2002) incorporated natural history data; and the third version combined RRMS and SPMS (Kobelt, Jönsson and Frederikson, 2003). All three compared a new biologic (interferon beta-1b, Betaferon®) to no treatment, as at that time none existed. The most recent model concerned a drug (natalizumab) that was introduced into a market where four treatments were well established and the comparison was therefore to existing treatments (Kobelt et al, 2008a).

The modelling approach is similar to the RA model in Study Example 8. The initial Markov models had six states based on a measure of functional disability (EDSS assessed by physicians) and one state for dead. The most recent model used ten states to capture the slow progression of MS better. In the absence of a relapse (see Study Example 1), EDSS scores correlate well with resource utilisation and quality of life, regardless of the type of MS or the country. Relapses cause additional costs and an additional disutility for their duration. Table 3.16 shows the mean annual costs and utilities by state in the most recent example.

Patients enter the model in a given state. At each cycle the model first verifies the probability of death, then the probability of having a relapse during the cycle, then the probability of having remained on treatment, and finally the development of EDSS, as schematically illustrated in Figure 3.24.

Figure 3.24. Structure of the Markov model



The first years in all models in this example are based on clinical trial data. The studies with interferon beta-1b included several hundred patients who were followed up during the trials for three to five years, regardless of whether they stayed on or discontinued treatment. Such trials are powerful enough to estimate transition probabilities between the EDSS states for patients with or without a relapse, as well as to assess treatment compliance. Also, as three-monthly measurements were available, a cycle length of three months was used in order not to lose any information. The analysis modelled the intervention for the duration of the trial only, as in the RA example with infliximab. Results of the three early models are summarised in Table 3.17 with differences in the estimates explained below.

Table 3.17. Cost per QALY with three-year treatment with interferon beta-1b in Sweden, all cost included

Ten-year model (costs and QALYS discounted 3%)	Incremental cost	QALY gain	Cost per QALY gained (SEK/ QALY)
SPMS, extrapolation based on clinical trial	55,500	0.162	342,600
SPMS, extrapolation based on natural history cohort	55,770	0.217	257,000
RRMS and SPMS, active patients extrapolation based on natural history cohort	13,700	0.207	66,200

Source: Kobelt et al (2000); Kobelt et al (2002); Kobelt, Jönsson and Frederikson (2003)

Initially, extrapolation beyond the trial was based on mean progression and relapse rates in the placebo group and discontinuation in the active group during the clinical trial. This is frequently done when no epidemiological data are available and the rates are then modified in a sensitivity analysis. Over a ten-year period, the cost per QALY gained with a three-year intervention in patients with SPMS was estimated at 342,600 SEK (Kobelt et al, 2000).

Subsequently, data for 824 patients with SPMS in a 30-year natural history database from Canada were incorporated into the model. When combining such datasets, it is important to assess the similarity of the patient populations. In this analysis, patients were similar in terms of age at diagnosis, disease duration, time to conversion from RRMS to SPMS, and EDSS score at conversion. Combining the two databases was hence not problematic, even less so as transition probabilities between states were calculated conditional upon these characteristics. When the natural history data were used to extrapolate beyond the clinical trial, the cost per QALY gained fell to 257,000 SEK (Kobelt et al, 2002).

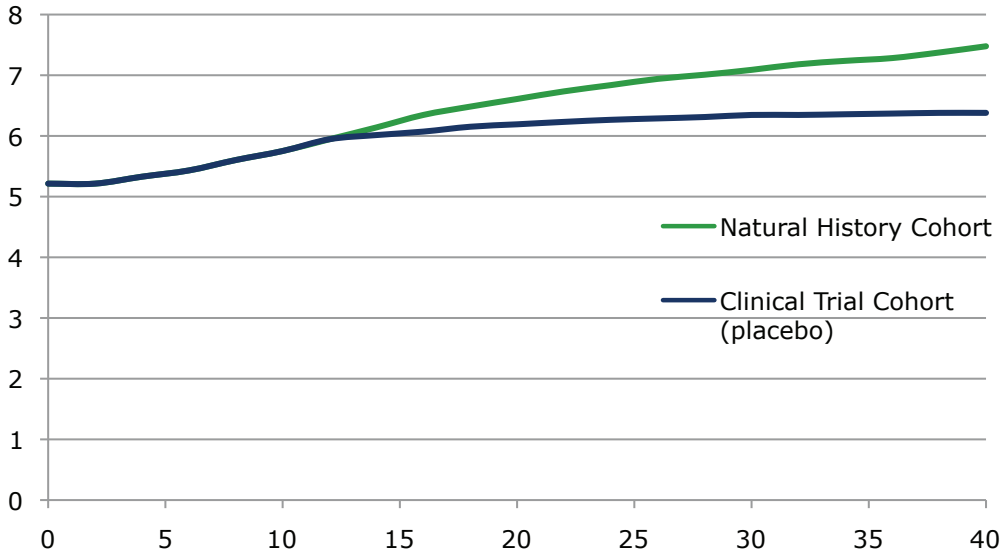
The reason for this improvement in the ICER is that extrapolation based on the clinical trial underestimated disease progression (see Figure 3.25). The main explanation for this was related to the type of patients enrolled in the trial, namely those with EDSS scores of between 3.0 and 5.5. Epidemiology has shown that patients will plateau for quite some time at an EDSS score of 6.0, before progression to 7.0 (wheelchair). As a consequence, a limited number of patients in a three-year trial will progress beyond 6.0; using these data for extrapolation will project this "plateau effect" forward and underestimate progression. A further important reason is the placebo effect in the clinical trial. Using the placebo group as the basis for extrapolation will logically also project this effect forward. This example underlines again the importance of the choice of datasets used to model diseases.

The third version of the model combined the two types of MS for patients with active disease and was based on two five-year trials with interferon beta-1b in patients with RRMS and SPMS and again used natural history data for extrapolation. Combining the two trials increases the number of observations, making the model more reliable and allowing confidence intervals and acceptability curves to be estimated. In this model, the cost per QALY gained is SEK 66,200 and the probability that the cost per QALY is less than SEK 500,000 for a patient starting in state 3 or 4 is 80% (Kobelt et al, 2003). The improvement of the ICER in this case is explained by the fact that the analysis was limited to patients with active disease and therefore at a higher risk of progression, and by treatment starting in earlier phases of the disease thereby leading to greater improvements in the long term.

Several issues need to be mentioned here. The fact that the cost per QALY decreases with each version of the model could potentially lead to doubts about the modelling process. However, the results are only changed due to the addition of more reliable data or looking at different groups of patients, not

due to the modelling process itself. Such a situation is quite common in fields where research is very active, or where high costs lead payers to limit access only to patients with the greatest need.

Figure 3.25. Extrapolation of disease progression



Source: Kobelt et al (2002)

Another issue is that treatment in the model is stopped at three (or five) years, and carries forward the benefit achieved. However, MS progression is very slow and, except when an exacerbation occurs, EDSS fluctuation is limited. This makes it more difficult to observe a benefit on progression or a loss of such a benefit after treatment discontinuation. The ICER thus expresses what can be achieved with treatment in terms of a lower EDSS score for however long treatment is given and, therefore, carrying this effect forward appears acceptable. Continuing treatment in the model beyond the period for which clinical data are available would involve making assumptions about the clinical effect at each level of disability, which is not a very good solution. This is different from the RA example, where the measure of disability is partly based on transient symptoms and the effect of treatment achieved during the trial therefore may not carry over fully.

In the more recent analysis of the cost-effectiveness of natalizumab (Kobelt et al, 2008a), the main issue was the comparator because the clinical trials were carried out against placebo rather than an active treatment. The model thus had to be updated to include current standard treatment, instead of no treatment. Rather than perform an indirect comparison to clinical trials using the standard drugs, the choice was made to use clinical practice data from the Swedish MS registry in Stockholm. At the time of data extraction, patients in the Stockholm area represented 42% (n=2,878) of all patients in the registry, of which almost half (n=1,316) had been or were being treated with one of the four existing treatments. These were matched with the trial patients and a final sample of 512 patients were eligible for use in the model. No distinction was made between the treatments. Instead, patients were defined as being on treatment, between treatments or off treatment. Treatment costs were calculated as the weighted average of treatments actually used.

Transition probabilities between the health states were then calculated in identical fashion for the sample from the registry and the sample from the natalizumab trial using a probit regression model. The resulting estimate of disease progression was compared to the original data for verification. For the simulation, both arms started with the cohort distribution in the active group of the natalizumab trial and simulations were run for 20 years, adopting a societal perspective with costs and effects discounted at 3%.

Total costs in the two arms were similar, although natalizumab offered a very small saving (<1% of total costs) despite its higher price. At the same time, natalizumab offers a slightly greater benefit,

which technically means that it dominated standard treatment. However, in such a case one would conclude that the two treatments are equivalent, first, because differences are minimal and, second, because the natalizumab data came from a controlled environment whereas standard treatment did not. The sensitivity analysis further illustrates this (Table 3.18), as the results are sensitive to very small changes in parameters, particularly changes in the effectiveness of natalizumab.

Table 3.18. Base case and sensitivity analysis, treatment with natalizumab in Sweden (2005 EUR)

Scenario	Total cost	Incre-mental cost	Total effect (QALY)	Incremental effect (QALY)	ICER (EUR/ QALY)
Base case (20 years, societal perspective, 3% discounted)					
Standard	613,680		8.99		
Natalizumab	609,850	-3,830	9.33	.034	dominant
Sensitivity analyses on natalizumab data					
Lower discontinuation rate					
• standard 5%	623,423		8.91		
• natalizumab 2.5%	647,839	24,416	0.34	0.43	56,811
• natalizumab 5.0%	618,647	-4,776	9.24	0.34	dominant
Reduced treatment effect					
• standard as observed	613,680		8.99		
• natalizumab -5%	621,455	7,775	9.25	0.26	30,275
• natalizumab -10%	633,210	19,530	9.16	0.17	113,450
Time horizon 10 years					
• standard	286,520		5.97		
• natalizumab	308,735	22,215	6.15	0.18	124,100

Source: Kobelt et al (2008a)

This particular study also illustrates a further issue that complicates the economic analysis. Due to a rare, but severe, adverse event, natalizumab was authorised only as second-line treatment for patients with active disease. This led to a situation wherein no direct efficacy data were available, as the trial had predominantly enrolled patients receiving first-line therapy. The results of the economic model imply, de facto, that the effect shown in the trial is the same in patients in whom a previous disease modifying treatment had failed. This may or may not be the case.

Cost-Benefit Analysis

As mentioned in the previous chapter, it is not always possible to reduce the outcomes of alternative treatment programmes to a single effect common to both alternatives. Cost-utility analysis offers one approach to solving this problem by using the QALY as a common measure of effectiveness.

An alternative method is cost-benefit analysis where both costs and outcomes are measured in monetary terms (Johannesson and Jönsson, 1991). With costs and benefits expressed in the same unit of measurement, it is possible to judge whether a project is desirable from a societal viewpoint, i.e. benefits are greater than costs. In addition, cost-benefit analysis enables the comparison of health care investments not only with other investments in the health care sector, but also with investments in non-health sectors such as education.

Few cost-benefit studies for health care interventions have been published, however. One reason is ethical objections to placing a monetary value on health, particularly with respect to valuing a human life. This is despite numerous everyday examples where health is valued in monetary terms, such as compensation for death and disability, and public expenditure on road safety projects. A second reason is that cost-minimisation, cost-effectiveness or cost-utility analysis will often yield sufficient data for resource allocation decisions to be made, and cost-benefit analysis is not needed.

In a cost-benefit analysis, a health care programme is considered good value for money when the value of the total benefit exceeds the total costs. Costs are ideally measured as opportunity costs, i.e. the best alternative benefit. Benefits are best measured by the maximum willingness to pay (WTP) for the outcomes of a project.

The theoretical basis for cost-benefit analysis is economic welfare theory and the concept of consumer surplus, i.e. WTP over and above the price actually paid, developed more than 50 years ago. The methods to measure health outcomes in monetary terms, however, have been adapted only recently. The standard method, contingent valuation, uses survey methods to measure individual WTP and was originally developed for valuing environmental benefits, for which it still is widely used.

Contingent valuation

Contingent valuation questions are either open-ended or discrete. In an open-ended valuation, the respondents are asked to state their maximum WTP for the benefit. The technique most used, "bidding game", resembles an auction, where a first bid is made to the respondent who then either accepts or rejects it. Depending on the answer, the bid is then lowered or increased until the respondent's maximum WTP is reached. In the alternative method, discrete questions of the yes/no or binary type are asked, which means that the respondent accepts or rejects the bid. Through varying the bid in different sub-samples, it is possible to calculate the percentage of respondents who are willing to pay as a function of price, i.e. the bid price.

As with all methods used in economic evaluation to value benefits, the contingent valuation method is well suited in some cases and not applicable in others. One situation in which the technique has shown good results is where the health gains can be well defined and where the patients know exactly what they are paying for, such as avoiding asthma attacks, angina attacks or episodes of pain. In an area such as prevention, although concerned with risk decisions of the type that individuals have to make in everyday life, the health gains are much more difficult to describe, and the probabilities of an event happening are usually small. This makes it more difficult for respondents to answer WTP questions. Currently, it appears that discrete binary questions provide better results, whereas in the bidding game the influence of the starting bid can heavily influence the results (for starting point bias see Study Example 10).

For the contingent valuation method to provide valid estimates of WTP, it must increase with the size of the health gain. This is clearly shown in the study examples below. The absolute figures obtained should, however, be interpreted with great caution. Even if individual WTP is related to the explanatory variables in the hypothesised way, it is still possible that the estimated WTP systematically underestimates or overestimates true individual WTP. Comparing hypothetical and true WTP for health changes is currently one of the important issues for research in this field.

In summary, cost-benefit analysis enables assessment of an individual's WTP for health gains by expressing the value of both the costs and benefits in monetary terms and thereby allows a comparative valuation of interventions across different sectors of the economy. An intervention is acceptable if the incremental benefits are greater than the incremental costs.

Study Example 10. Cost-benefit analysis and willlingness to pay—angina pectoris

WTP studies are best suited to disease areas where a patient-related benefit can be easily expressed as a single outcome measure. In the cardiovascular field, the contingent valuation method was used to assess individuals' WTP for a treatment that would reduce the number of angina attacks (Kartman, Andersson and Johannesson, 1996).

Angina pectoris is a widespread cardiovascular disease that is characterised by chest pain associated with transient episodes of myocardial ischaemia resulting from an imbalance between oxygen supply and tissue demand. There is no clear consensus on whether the severity of attacks can be defined by the degree of "pain" or "discomfort", although the frequency of attacks can be used to measure the severity of the disease.

The question put to 400 Swedish patients with angina pectoris was as follows: "Imagine that there are two treatments for your disease. The first is your current treatment; the second is more effective and has been shown to reduce weekly attacks by 50%. However, for each three-month period of the second treatment you have to pay a certain amount from your own income". (The percentage reduction in the frequency of attacks was varied between 25% and 75% in randomised subsamples)

The study used both the binary question and the bidding game techniques. The main problem with the bidding game approach is that the reported WTP is likely to be affected by the size of the first bid offered, a phenomenon called "starting point bias". With binary questions, each respondent only accepts or rejects one bid, and the bid is varied in different subsamples to determine the mean WTP for the group.

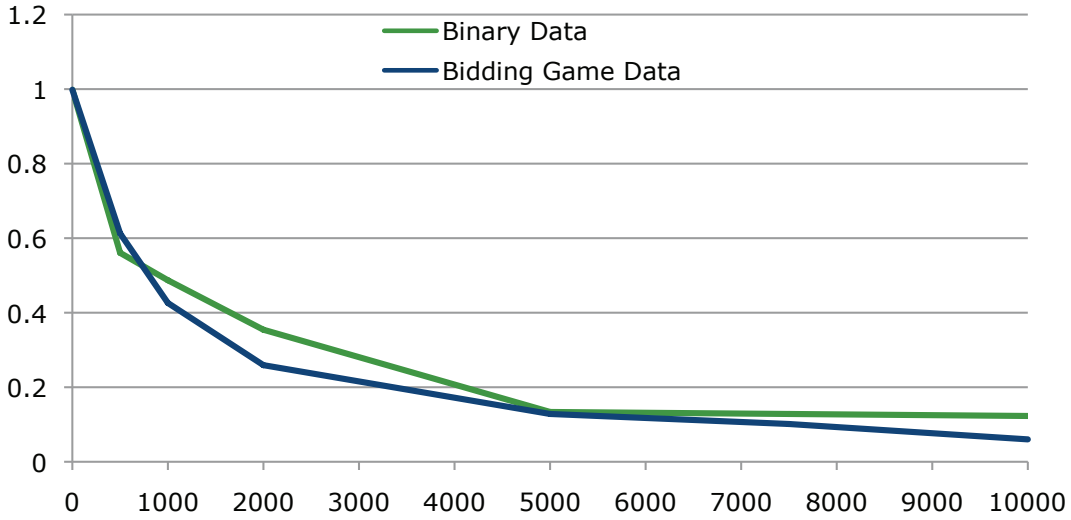
The answers were analysed using multiple regression models that included a set of explanatory variables capturing angina status, weekly attack rate and income levels of the respondents. It was hypothesised that WTP would rise with increasing severity of angina and an increasing weekly attack rate. Results are shown in Table 3.19. Figure 3.22 shows the proportion of individuals, for both the binary and the bidding game, that is willing to pay as a function of the bid.

Table 3.19. Mean WTP for different rates of reduction of angina attacks (1994 SEK)

Method	Attack rate reduction		
	25%	50%	75%
Binary question data	1,873	2,499	2,692
Bidding game data	1,388	2,079	3,350

Source: Kartman, Andersson and Johannesson (1996)

Figure 3.26. Proportion of individuals' WTP as a function of money (SEK). Pooled data for all attack rate reductions.



Source: Kartman et al (1996)

Study Example 11. Cost-benefit analysis and willingness to pay—incontinence

A very similar study investigated the WTP for a reduction in symptoms of urge incontinence, a condition in which the outcome is somewhat more difficult to express as a single patient-related benefit (Johannesson et al, 1997). Patients with urge incontinence experience symptoms of urgency, urinary frequency and involuntary loss of urine. Quality of life is impaired as urgency is often associated with colic-like pain and daytime urinary frequency can severely limit activities, while nocturnal frequency can be associated with persistent fatigue. Treatments include physiotherapy, pharmacological treatment and, in rare cases, surgery. At the time of this study, drug therapy was hampered by limited efficacy or severe side effects leading to extremely poor compliance. In the absence of cure or effective treatment, sanitary protections are widely used.

Frequency of micturitions and episodes of involuntary urine loss are not independent, as patients cope by making frequent visits to the bathroom to avoid episodes of leakage. Clinical trials will measure both symptoms separately, but for a WTP questionnaire it is important to express the outcome with one measure. This study investigated the possibility of combining these symptoms into one outcome measure and tested the appropriateness of the measure by assessing its correlation with health-related quality of life.

A specific WTP questionnaire, as well as EQ-5D and SF-36 questionnaires, were mailed to a sample of patients. The combined outcome measure was considered acceptable because it correlated significantly with all domains of the SF-36, as well as with EQ-5D utilities (Table 3.20).

Table 3.20 Quality of life of patients with urge incontinence and correlations between quality of life and symptoms

	Urge incontinence	Matched normals		Correlation coefficient	
	Mean (SD)	Mean (SD)	P	(Symptom score and QoL)	P
SF-36					
Physical functioning	66.0 (25.2)	75.3 (11.8)	<0.001	-0.22	<0.001
Role, physical	55.3 (43.0)	70.2 (14.1)	<0.001	-0.16	<0.001
Bodily pain	55.9 (26.9)	67.8 (4.8)	<0.001	-0.14	0.004
General health	56.3 (24.4)	67.7 (7.5)	<0.001	-0.23	<0.001
Vitality	53.7 (26.3)	64.6 (7.1)	<0.001	-0.19	<0.001
Social functioning	75.8 (26.0)	85.0 (4.6)	<0.001	-0.23	<0.001
Role, emotional	67.0 (40.9)	78.4 (9.4)	<0.001	-0.10	0.044
Mental health	70.5 (22.9)	78.3 (3.7)	<0.001	-0.17	0.001
EuroQol					
EQ-5D	0.68	0.80	<0.0001	-0.25	<0.001
Rating scale	65.56	79.0	<0.0001	-0.20	<0.001

Source: Adapted from Johannesson et al (1997)

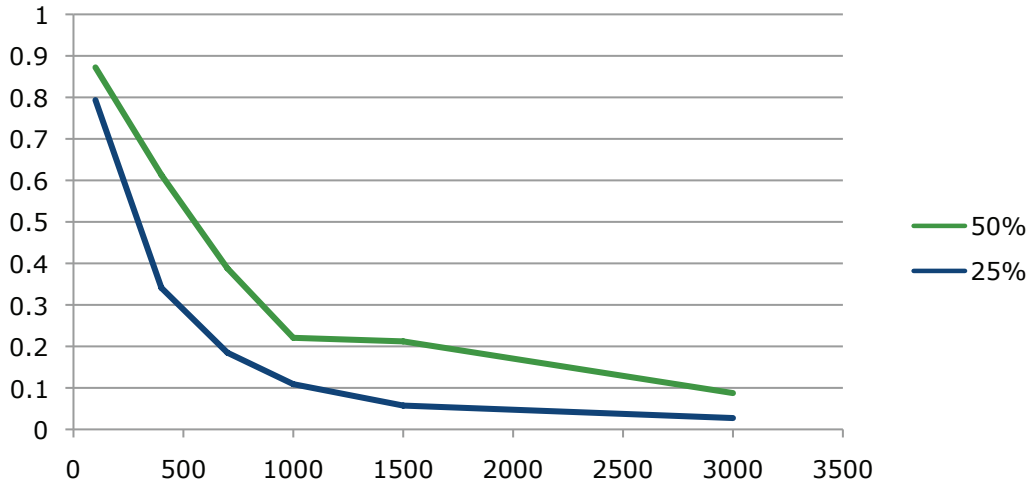
The WTP question was framed using the binary technique; patients were asked whether they would pay a given price for a given reduction in symptoms. The percentage reduction in symptoms was varied between 25% and 50% and six price levels were used. The range of prices, as well as the understanding of the question by the respondent, had been pre-tested. Patients were willing to pay more for the larger percentage reduction in symptoms; this one way of judging where respondents understand the WTP question (Table 3.21 and Figure 3.23). WTP increased with the severity of the symptoms and therefore the potential absolute benefit increased. It also increased with higher income, as expected (Table 3.22). Overall, this study is a very clear example of a WTP study where all parameters "behaved" as they should.

Table 3.21. Mean and median WTP for a reduction of incontinence symptoms

Reduction in the frequency of micturitions and episodes of incontinence		
	25%	50%
Median WTP	240 SEK	466 SEK
Mean WTP	529 SEK	1,027 SEK

Source: Johannesson et al (1997)

Figure 3.27. Proportion of patients' WTP for a reduction in symptoms as a function of the price of treatment



Source: Johannesson et al (1997)

Table 3.22. Sensitivity analysis: median and mean WTP for a 25% reduction in symptoms at different levels of symptom severity and income (in SEK)

Monthly income SEK	Median (mean) willingness to pay for varying varying levels of symptom severity		
	15	20	25
5,000	134 (294)	168 (409)	234 (515)
10,000	273 (601)	378 (832)	476 (1,047)
15,000	379 (834)	525 (1,155)	661 (1,545)

Source: Johannesson et al (1997)

Chapter 4

GUIDELINES FOR ECONOMIC EVALUATION

As discussed in the first chapter, a number of countries around the world in the past 20 years have formally incorporated an economic criterion into the decision making process for health care, principally when assessing whether a new pharmaceutical or medical device should be listed on a publicly-funded formulary. Each of these initiatives has been accompanied by the development of methodological guidelines designed specifically to support the particular policy requirement in the respective country.

The document from the Canadian Agency for Drugs and Technologies in Health (CADTH) may be seen as a generic set of guidelines for those undertaking economic evaluations. It is intended to be applicable to any health technology, despite the title, *Guidelines for economic evaluation of pharmaceuticals: Canada.* The original guidelines, issued in 1995, have been updated (CADTH, 2006) and complemented with specific recommendations relating to oncology products (CADTH, 2009).

In view of the fact that such detailed methodological guidelines have been available for so long, countries where economic evaluations have subsequently been made mandatory have often adopted a more general approach. An example are the guidelines published by the Swedish Dental and Pharmaceutical Benefits Board (TLV) in 2003. These leave aside detailed methodological recommendations and focus instead on the important concepts that may or may not be different from other guidelines (LFNAR, 2003). This document has been complemented by guidelines on how to structure submissions overall.

A more context-specific set of guidelines is the guidance for manufacturers and sponsors in making submissions to the UK's National Institute for Health and Care Excellence (NICE). The initial 1999 guidelines were revised in 2004, 2008 and, most recently, in 2013 (NICE, 2013a). The original plan was for NICE to review several technologies or indications as part of multiple technology appraisals (MTA), as is usual for HTA, but in 2007 NICE introduced single technology appraisals (STA) to allow faster assessment of novel technologies and medicines. The NICE guidance, unlike the CADTH or TLV documents, was developed to support a particular programme of technology appraisal in England and Wales. As a result, the guidance is more detailed with regards to the specific methods requested and more stringent as far as the setting of the analysis is concerned, including a reference case defining the format and content of submissions from manufacturers.

The three sets of guidelines illustrate well the differences in approach by countries or authorities. Key among these are the appropriate perspective for the analysis, the type of economic evaluation, the acceptance of studies from other settings and the approach to discounting.

1. *Perspective:* In accordance with the methodological literature generally, the CADTH guidelines strongly recommend that a societal perspective be adopted, but request that results be presented from other viewpoints, including that of the primary decision maker. All costs, no matter who incurs them, are considered relevant. Sweden's TLV guidelines request specifically that a societal perspective be adopted, including costs in added years of life. Analyses thus also include costs falling on sectors of the economy outside health care, e.g. those borne by the patient and family and, where appropriate, time costs (which include informal care and production losses). The NICE guidance, on the other hand, focuses exclusively on costs to the NHS and Personal Social Services (PSS), excluding all other costs.

2. *Type of study:* As far as the type of economic evaluation is concerned, the NICE guidance recommends the use of cost-utility analysis, whereas the CADTH and TLV guidelines take a broader view of the type of analysis admissible. However, both CADTH and TLV express

a clear preference for cost-utility analysis or cost-benefit analysis. Where cost-utility analysis is undertaken, the three guidelines agree that health state preferences of the public should be used. However, while CADTH and TLV are relatively flexible as to the methods and the population used to elicit utilities, NICE demands that utility be measured using a choice-based method and considers the preferences of the general population in England and Wales as the most relevant for submissions.

3. *Transfer of economic studies:* The TLV guidelines, as is typical for recommendations in smaller countries, explicitly accept studies that are transferred from other countries, with appropriate adaptation. The same attitude is implicit in the CADTH guidelines. In contrast, NICE expects that the study be undertaken in England and Wales.

4. *Discounting:* While in general a discount rate of 3% is applied in most countries (including Sweden), Canada maintains a rate of 5%. Both guidelines discount costs and health benefits with the same discount rate, but Sweden also requests analyses without discounting and with only costs discounted. This reflects ongoing discussion of whether outcomes should be discounted in the same way as costs. This debate is reflected in NICE guidance: in the first set of guidelines, costs were discounted at 6% and benefits at 1.5%; in later editions, a rate of 3.5% for both costs and benefits was adopted. Although often viewed as a minor technicality, the choice of discount rates and how they are applied can have a substantial impact on cost-effectiveness results, depending on the time frame.

Most guidelines have been developed with the aid of a review of other guidelines. Areas of common ground in methods advocated by different organisations are substantial as a result. Recent and continuing efforts by European HTA agencies are intended to produce enough methodological agreement to allow partial sharing of analyses (EUnetHTA, 2013). Similarities already are obvious: most guidelines emphasize the need for effectiveness data, rather than relying on efficacy measures, and recognise the importance of modelling techniques to translate one into the other. Most focus on existing practice or the most-used alternative as the relevant comparator. Most regard subgroup analyses as admissible, with the proviso that they are based on prior reasoning, if possible. All sets of recommendations prompt analysts to allow for equity concerns by identifying those groups of patients that would be most likely to benefit from the intervention being evaluated. Various other points of consensus across countries are evident. In practice, then, a core set of agreed methodological principles already exists, but with individual countries adjusting them to fit their own particular requirements.

Use of guidelines, long confined to advanced economies, is expanding. Emerging market countries currently are creating guidelines, particularly in Latin America and Asia, and developing approaches for their decision making. Although the guidelines will closely echo existing guidelines, the process of creating the guidelines can itself be both an educational process and a means for encouraging consensus among stakeholders.

Chapter 5

CONCLUSIONS

Health economics applies the theories, tools and concepts of economics to health and health care. It is now a central tool in health policy makers' efforts to introduce more efficiency in health care organisation, financing and resource allocation. Policy makers, HTA agencies, payers, providers and patients are all involved. The question is no longer if, but how, economic evaluations can best be used in health care.

Economic evaluations analyse the consequences, in terms of costs and benefits, of using new or established therapies compared to available alternatives. They provide part of the basis for making decisions about resource allocation within the confines of limited budgets. The basic methodology for economic evaluation has been in use for some time, but it will continue to be subjected to research and scrutiny as the field continues to evolve. Important issues will be revisited over time as recommendations for the conduct of economic evaluation are debated and broader groups of stakeholders provide input. State-of-the art methodological guidelines, such as the CADTH document, and other publications on good methodological principles, including this book, should therefore be seen as living documents and as works in progress.

The practical examples in this book have focused largely on the economic evaluation of pharmaceuticals, the focus of most policy attention to date on the use of an economic criterion. However, many of the issues covered are equally relevant for non-pharmaceutical technologies. Indeed, efforts to assess medical devices or interventional techniques in the same way as pharmaceuticals are becoming more common, and the methods that have been developed are well suited to this.

Technologies other than drugs present their own sets of challenges. One is the dearth of clinical data from trials that are often too small and too short to provide enough evidence of the outcome; and for certain devices and procedures it is difficult or infeasible to undertake a randomised trial. In addition, outcomes for some technologies may depend as much, or more, on the skills of the professionals involved, such as surgeons, than on the technology per se.

Cost-effectiveness studies of pharmaceuticals are often performed prior to their introduction to aid in decisions as to whether to include them in reimbursement schemes, and if so, for which patient populations and at what price. Such studies represent a specific aspect of HTA that deals with the introduction of new technologies. The evaluation of the use of existing technologies and methods in clinical practice involves a larger group of stakeholders and use of "real world" and patient registry data.

Each of the examples in this book illustrates a particular method or issue, but all involve modelling that combines data from different sources. Modelling is almost always necessary, as it is very rare that one data set delivers all the necessary information. Modelling techniques are widely used and accepted and are likely to remain so, particularly for submissions to reimbursement authorities by pharmaceutical companies before marketing, when data on costs and effects is very limited. Acceptance of models before marketing begins, however, may be accompanied by a legitimate demand for follow-up studies in actual clinical practice to verify the results of the earlier modelling studies.

Despite the interest in follow-up studies, experience with using clinical practice data in populating models is limited, as several difficulties arise. First, the first patients to receive a new treatment are often at the severe end of the disease spectrum, with the new treatment seen as a "last resort" and thus not necessarily comparable to the scenarios modelled. Second, it can take many years to collect

relevant data for such verifications, particularly in chronic diseases; data collected routinely in registries are not always adequate for economic evaluation. Finally, it is challenging to design follow-up studies that fulfil the requirements for both outcome and economic analyses. An example of this difficulty is the much-discussed risk-sharing agreement for MS drugs in the UK (Sudlow and Counsell, 2003). Nevertheless, the need for follow-up studies is one of the two areas where changes are most likely in the coming years; the other is increasing demand for comparative effectiveness data.

As economic evaluations are presented to increasingly wider audiences, and used as tools by a wider range of decision makers, the issue of study quality acquires greater significance. To familiarise the medical professional audience with these issues, studies increasingly are being published in peer-reviewed medical journals, rather than only health economics journal. This presents a challenge for reviewers, whose medical backgrounds may not be sufficient for assessing economic analyses. The British Medical Journal approached this issue by publishing a checklist for the critical review of economic evaluations (Drummond and Jefferson, 1996). While this checklist approach (see Figure 5.1) cannot ensure thorough, critical appraisals of economic evaluations, it certainly can provide a valuable guide for reviewing, and writing, study reports and papers.

As the search for greater efficiency and value for money in health care continues, economic analyses of health care technologies will find a ready audience among decision makers of all types. At the same time, refinements to the techniques of economic analysis will continue to improve current methods and also address decision makers' evolving concerns. New complexities and challenges can be expected to arise as, for example, technology increasingly allows treatment to be targeted more precisely to the individual's genetic make-up.

Figure 5.1. British Medical Journal's checklist for reviewers or referees

Item	Yes	No	Not Clear	Not appropriate
Study design				
1. The research question is stated	❑	❑	❑	
2. The economic importance of the research question is stated	❑	❑	❑	
3. The viewpoint(s) of the analysis are clearly stated and justified	❑	❑	❑	
4. The rationale for choosing the alternative programmes or interventions compared is stated	❑	❑	❑	
5. The alternatives being compared are clearly described	❑	❑	❑	
6. The form of economic evaluation used is stated	❑	❑	❑	
7. The choice of form of economic evaluation is justified in relation to the questions addressed	❑	❑	❑	
Data collection				
8. The source(s) of effectiveness estimates used are stated	❑	❑	❑	
9. Details of the design and results of effectiveness study are given (if based on a single study)	❑	❑	❑	❑
10. Details of the method of synthesis or meta-analysis of estimates are given (if based on an overview of a number of effectiveness studies)	❑	❑	❑	❑
11. The primary outcome measure(s) for the economic evaluation are clearly stated	❑	❑	❑	

Item	Yes	No	Not Clear	Not appropriate
12. Methods to value health states and other benefits are stated	❑	❑	❑	❑
13. Details of the subjects from whom valuations were obtained are given	❑	❑	❑	❑
14. Productivity changes (if included) are reported separately	❑	❑	❑	❑
15. The relevance of productivity changes to the study question is discussed	❑	❑	❑	❑
Study design				
16. Quantities of resources are reported separately from their unit costs	❑	❑	❑	
17. Methods for the estimation of quantities and unit costs are described	❑	❑	❑	
18. Currency and price data are recorded	❑	❑	❑	
19. Details of currency of price adjustments for inflation or currency conversion are given	❑	❑	❑	
20. Details of any model used are given	❑	❑	❑	❑
21. The choice of model used and the key parameters on which it is based are justified	❑	❑	❑	❑
Analysis and interpretation of results				
22. Time horizon of costs and benefits is stated	❑	❑	❑	
23. The discount rate(s) is stated	❑	❑	❑	❑
24. The choice of rate(s) is justified	❑	❑	❑	❑
25. An explanation is given if costs or benefits are not discounted	❑	❑	❑	❑
26. Details of statistical tests and confidence intervals are given for stochastic data	❑	❑	❑	❑
27. The approach to sensitivity analysis is given	❑	❑	❑	❑
28. The choice of variables for sensitivity analysis is justified	❑	❑	❑	❑
29. The ranges over which the variables are varied is stated	❑	❑	❑	❑
30. Relevant alternatives are compared	❑	❑	❑	❑
31. Incremental analysis is reported	❑	❑	❑	❑
32. Major outcomes are presented in a disaggregated as well as aggregated form	❑	❑	❑	
33. The answer to the study question is given	❑	❑	❑	

Source: Drummond and Jefferson (1996)

GLOSSARY

Average cost	Total cost of therapy divided by the total quantity of treatment units provided
Bayesian analysis	An approach to statistical analysis that allows prior evidence and beliefs to be incorporated formally into the analysis of new data
Bootstrapping	A technique that involves resampling with replacement of patient data from an existing data set. By performing multiple repetitions (1,000 or more) of this procedure, simulated distributions of variables of interest, such as mean costs, mean effects and the incremental cost-effectiveness ratio (ICER) can be derived. Uncertainty around these statistics can then be explored without making assumptions about their distribution.
Burden/cost of illness study	A descriptive study that relates direct and indirect costs to a defined illness
Confidence interval	A range of values that contains the true value of the variable of interest a given percentage (e.g. 95%) of the time in repeated sampling
Contingent valuation	A method of eliciting individuals' preferences for a service by asking how much they are hypothetically willing to pay for the service. It is the technique conventionally used to obtain and attach monetary values to the benefits of health care in cost-benefit analysis
Cost-benefit analysis	Type of economic evaluation that measures costs and benefits in monetary units and computes a net pecuniary gain/loss
Cost-effectiveness	Efficient use of (scarce) resources
Cost-effectiveness acceptability curve	A line showing the proportion of estimates of the ICER falling below the threshold ICER for different values of the threshold, frequently interpreted as the probability that the intervention is cost-effective
Cost-effectiveness analysis	Type of economic evaluation that measures therapeutic effects in physical or natural units and computes a cost:effectiveness ratio for comparison purposes
Cost-minimization analysis	Type of economic evaluation that finds the lowest cost programme among those shown to be of equal benefit
Cost-utility analysis	Type of analysis that measures therapeutic consequences in utility units (e.g. QALYs) rather than in physical units
DALY	The disability-adjusted life year, a measure akin to the QALY in aggregating survival and quality of life effects, but normally advanced as a method of estimating the burden of illness associated with a disease, rather than the cost-effectiveness of health care interventions

Decision analysis	An explicit quantitative approach to decision making under uncertainty, with a structure designed to represent the treatment options under investigation and normally based on a synthesis of data from the literature
Direct medical costs	Fixed and variable costs associated directly with a health care intervention
Direct non-medical costs	Non-medical costs associated with provision of medical services
Discounting	The adjustment of future costs and benefits to render those occurring in different years comparable with each other and with current costs and benefits. The adjustment operates in the opposite way to compound interest, i.e. a positive discount rate weights the future less than the present.
Disease management	A health care management process bringing together the development and delivery of all health care interventions and costs relevant to the prevention and management of a particular disease
Economic evaluation	A comparative analysis of two or more alternatives in terms of their costs and consequences
Effectiveness	The therapeutic consequence of a treatment in real world conditions, for example in a clinical trial
Efficacy	The consequence (benefit) of a treatment under ideal and controlled clinical conditions, for example in a clinical trial
Health economics	Application of the theories, concepts and tools of economics to health and health care
Health-related quality of life (HRQoL)	The impact on an individual's well-being of their health, often encompassing physical, mental and psychosocial elements
Health state	A summary description of an individual's health-related quality of life
Health technology assessment (HTA)	A multidisciplinary process that summarises information about the medical, social, economic and ethical issues related to the use of a health technology in a systematic, transparent, unbiased, robust manner
HYE	The healthy-years equivalent, a summary measure of health outcome analogous to the QALY in combining survival with quality of life, derived using a two-stage standard gamble technique
Incremental cost	The additional cost that one service or programme imposes over another, mutually-exclusive alternative
Incremental cost-effectiveness ratio (ICER)	The additional cost of producing an extra unit of outcome by one therapy compared with another
Indirect comparison	A statistical comparison of alternatives that have not been compared head-to-head; also called "network meta-analysis"
Indirect costs or productivity costs	Cost of reduced productivity resulting from illness or treatment

Intangible costs	The cost of pain and suffering as a result of illness or treatment
Marginal cost	The extra cost of one extra unit of product or service delivered
Markov analysis	A modelling technique to handle decision problems involving risks that are potentially continuously variable over time and where the timing of the events is important
Meta-analysis	A systematic process for finding, evaluating and combining the sets of data from different scientific studies and combining the results
Moral hazard	A change in behaviour of buyers or sellers as a result of insurance. Insurance changes behaviour because it alters the level of financial risk faced by buyers and sellers.
Net benefit (NB)	A summary measure of the difference between an intervention's mean incremental health effects (AE, normally measured in QALYs) and its mean incremental costs (AC) relative to an alternative. The incremental net benefit (NB) can be expressed in monetary terms (the money value of AE minus AC) or, less frequently, health terms A positive NB implies that the ICER is within the threshold ICER.
Network meta-analysis	A statistical method that enables a comparison of two or more products when head-to-head data are lacking; also called "indirect comparison"
Opportunity cost	The benefit foregone from using a resource for one purpose as opposed to its best alternative use
Outcomes research	The study of the ultimate therapeutic consequences of a treatment, including its effect on patients' quality of life
Pharmacoeconomics	The economic evaluation of pharmaceutical products
Probabilistic sensitivity analysis	A technique used to explore the impact on a simulated group of patients (such as those entered into a Markov model) of uncertainty around estimates of the input parameters
QALY	The quality-adjusted life year is the outcome of a treatment measured as the number of years of life saved, adjusted for their utility (quality).
Sensitivity analysis	The assessment of the robustness of study results through systematic variation of key variables
Standard gamble (SG)	A method of valuing health states on a 0–1 scale by presenting individuals with a choice between a given health state for certain and a gamble offering (for better than death states) outcomes of death (valued as 0) and perfect health (1). The probability of perfect health at which the individual would be indifferent between the two options gives the value of the health state.
Threshold ICER	The maximum willingness to pay for health benefits, normally expressed as the maximum cost per QALY that decision makers consider acceptable for a health care intervention

Time trade off (TTO)	A means of valuing health states on a 0–1 scale by asking individuals how many years in perfect health are equivalent to a given number of years in a less than perfect health state. Years in perfect health divided by years in the defined health state gives the value for that health state.
Utility	A measure of the relative preference for, or desirability of, a specific level of health status or a specific health outcome
Visual analogue scale (VAS)	A means of valuing health states on a 0–1 scale by asking individuals individuals to place them on a line ranging from best possible health (valued as 1) to worst possible health or death (valued as 0)

REFERENCES

Arikian, S.R., Einarson, T.R., Kobelt-Nguyen, G. and Schubert, F. (1994) A multinational pharmacoeconomic analysis of oral therapies for onychomycosis. The Onychomycosis Study Group. *British Journal of Dermatology.* 130(Suppl43), 35-44.

Briggs, A. (2000) Economic evaluation and clinical trials: Size matters. *British Medical Journal.* 321(7273), 1362-1363.

Briggs, A.H., Wonderling, D.E. and Mooney, C.Z. (1997) Pulling cost-effectiveness analysis up by its bootstraps: A non-parametric approach to confidence interval estimation. *Health Economics.* 6(4), 327-340.

CADTH (Canadian Agency for Drugs and Technologies in Health). (2006) *Guidelines for the economic evaluation of health technologies: Canada.* 3rd Edition. Ottawa: CADTH.

CADTH (Canadian Agency for Drugs and Technologies in Health). (2009) *Addendum to CADTH's guidelines for the economic evaluation of health technologies: Specific guidance for oncology products*. Ottawa: CADTH.

Cleland, J., Takala, A., Apajasalo, M., Zethraeus, N. and Kobelt, G. (2001) PCV11: The cost-effectiveness of levosimendan in severe low-output heart failure compared to dobutamine based on an international clinical trial (LIDO). *Value in Health.* 4(6), 492.

Dahlberg, L., Lundkvist, J. and Lindman, H. (2009) Health care costs for treatment of disseminated breast cancer. *European Journal of Cancer.* 45(11), 1987-1991.

Devlin, N. and Krabbe, P. (2013) The development of new research methods for the valuation of EQ-5D-5L. *European Journal of Health Economics.* 14(1Suppl), 1-3.

Dolan, P., Gudex, C., Kind, P. and Williams, A. (1995) *A social tariff for EuroQol: Results from a UK general population survey*. Working paper 138. York: University of York Centre for Health Economics.

Drummond, M.F. and Jefferson, T.O. (1996) Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *British Medical Journal.* 313(7052), 275-283.

Drummond, M.F., Sculpher, M.J., Torrance, G.W., O'Brien, B.J., and Stoddart, G.L. (2005) *Methods for the economic evaluation of health care programmes.* Third edition. New York: Oxford University Press.

Elliott, W.J. and Meyer, P.M. (2007) Incident diabetes in clinical trials of antihypertensive drugs: A network meta-analysis. *The Lancet.* 369(9557), 201-207.

EUnetHTA. (2013). EUnetHTA. Available at http://www.eunethta.eu/. [Accessed 2 August 2013].

Fieller, E.C. (1954) Some problems in interval estimation. *Journal of the Royal Statistical Society. Series B (Methodological).* 16(2), 175-185.

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

Hardens, M., Vos, W.F.H., Kobelt-Nguyen, G. and Souetre, E. (1995) *A retrospective economic analysis of Sandimmun Neoral in de novo and in stable renal transplantation patients in Germany, Austria, Switzerland and Italy.* Paper presented at the European Symposium on Pharmacoeconomics. European Society of Clinical Pharmacy.

Henriksson, F., Fredrikson, S., Masterman, T. and Jönsson, B. (2001) Costs, quality of life and disease severity in multiple sclerosis: A cross-sectional study in Sweden. *European Journal of Neurology.* 8(1), 27-35.

Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., Bonsel, G. and Badia, X. (2011) Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research.* 20(10), 1727-1736.

Higgins, J.P.T. and Green, S. eds. (2011) Cochrane handbook for systematic review of interventions. Version 5.1.0. Available at http://handbook.cochrane.org/ [Accessed 1 August 2013].

Johannesson, M. and Jönsson, B. (1991) Economic evaluation in health care: Is there a role for cost-benefit analysis? *Health Policy.* 17(1), 1-23.

Johannesson, M., Jönsson, B., Jönsson, L., Kobelt, G. and Zethraeus, N. (2009) Why should economic evaluations of medical innovations have a societal perspective? *Briefing.* No. 51. London: Office of Health Economics.

Johannesson, M., Meltzer, D. and O'Conor, R.M. (1997) Incorporating future costs in medical cost-effectiveness analysis: Implications for the cost-effectiveness of the treatment of hypertension. *Medical Decision Making.* 17(4), 382-389.

Johannesson, M., O'Conor, R.M., Kobelt-Nguyen, G. and Mattiasson, A. (1997) Willingness to pay for reduced incontinence symptoms. *British Journal of Urology.* 80(4), 557-562.

Jönsson, B. (2009) Ten arguments for a societal perspective in the economic evaluation of medical innovations. *European Journal of Health Economics.* 10(4), 357-359.

Jönsson, B. and Karlsson, G. (1990) Economic evaluation of cancer treatments. *In* Domellöf, L. ed. *Drug delivery in cancer treatment III.* Berlin Heidelberg: Springer-Verlag. 63-84.

Jönsson, B. and Krieglstein, G.K. (1998) *Primary open-angle glaucoma: Differences in international treatment patterns and costs.* Oxford: ISIS Medical Media.

Kartman, B., Andersson, F. and Johannesson, M. (1996) Willingness to pay for reductions in angina pectoris attacks. *Medical Decision Making.* 16(3), 248-253.

Kobelt, G. and Jönsson, L. (1999) Modeling cost of treatment with new topical treatments for glaucoma. Results from France and the United Kingdom. *International Journal of Technology Assessment in Health Care.* 15(1), 207-219.

Kobelt, G., Berg, J., Lindgren, P., Elias, W.G., Flachenecker, P., Freidel, K., Koenig, N., Limmorth, V. and Straube, E. (2006a) Costs and quality of life of multiple sclerosis in Germany. *European Journal of Health Economics.* 7(2Suppl), 34-44.

Kobelt, G., Berg, J., Lindgren, P., Fredrikson, S. and Jönsson, B. (2006b) Costs and quality of life of patients with multiple sclerosis in Europe. *Journal of Neurology, Neurosurgery & Psychiatry.* 77(8), 918-926.

Kobelt, G., Berg, J., Lindgren, P., Jönsson, B., Stawiarz, L. and Hillert, J. (2008a) Modeling the cost-effectiveness of a new treatment for MS (natalizumab) compared with current standard practice in Sweden. *Multiple Sclerosis.* 14(5), 679-690.

Kobelt, G., Eberhardt, K., Jönsson, L. and Jönsson, B. (1999) Economic consequences of the progression of rheumatoid arthritis in Sweden. *Arthritis & Rheumatism.* 42(2), 347-356.

Kobelt, G., Jönsson, L. and Fredrikson, S. (2003) Cost-utility of interferon beta1b in the treatment of patients with active relapsing-remitting or secondary progressive multiple sclerosis. *European Journal of Health Economics.* 4(1), 50-59.

Kobelt, G., Jönsson, L., Henriksson, F., Fredrikson, S. and Jönsson, B. (2000) Cost-utility analysis of interferon beta-1b in secondary progressive multiple sclerosis. *International Journal of Technology Assessment in Health Care.* 16(3), 768-780.

Kobelt, G., Jönsson, L., Lindgren, P., Young, A. and Eberhardt, K. (2002) Modeling the progression of rheumatoid arthritis: A two-country model to estimate costs and consequences of rheumatoid arthritis. *Arthritis & Rheumatism.* 46(9), 2310-2319.

Kobelt, G., Jönsson, L., Young, A. and Eberhardt, K. (2003) The cost-effectiveness of infliximab (Remicade) in the treatment of rheumatoid arthritis in Sweden and the United Kingdom based on the ATTRACT study. *Rheumatology.* 42 (2), 326-335.

Kobelt, G., Lindgren, P., Lindroth, Y., Jacobson, L. and Eberhardt, K. (2005a) Modelling the effect of function and disease activity on costs and quality of life in rheumatoid arthritis. *Rheumatology.* 44(9), 1169-1175.

Kobelt, G., Lindgren, P., Singh, A. and Klareskog, L. (2005b) Cost effectiveness of etanercept (Enbrel) in combination with methotrexate in the treatment of active rheumatoid arthritis based on the TEMPO trial. *Annals of the Rheumatic Diseases.* 64(8), 1174-1179.

Kobelt, G., Lindgren, P., Smala, A., Bitsch, A., Haupts, M., Kölmel, H.W., König, N., Rieckmann, P. and Zettl, U.K. (2001) Costs and quality of life in multiple sclerosis. *European Journal of Health Economics.* 2(2), 60-68.

Kobelt, G., Woronoff, A.S., Richard, B., Peeters, P. and Sany, J. (2008b) Disease status, costs and quality of life of patients with rheumatoid arthritis in France: The ECO-PR Study. *Joint Bone Spine.* 75(4), 408-415.

Lekander, I., Kobelt, G., Svarvar, P., Ljung, T., van Vollenhoven, R. and Borgström, F. (2013) The comparison of trial data-based and registry data-based cost-effectiveness of infliximab treatment for rheumatoid arthritis in Sweden using a modeling approach. *Value in Health.* 16(2), 251-258.

Lewis, S. and Clarke, M. (2001) Forest plots: Trying to see the wood and the trees. *British Medical Journal.* 322(7300), 1479-1480.

LFNAR (Pharmaceutical Benefits Board). (2003) General guidelines for economic evaluations from the Pharmaceutical Benefits Board. Available at http://www.tlv.se/Upload/English/Guidelines-for-economic-evaluations-LFNAR-2003-2.pdf. [Accessed 13 August 2013].

Lindgren, P., Geborek, P. and Kobelt, G. (2009) Modeling the cost-effectiveness of treatment of rheumatoid arthritis with rituximab using registry data from Southern Sweden. *International Journal of Technology Assessment in Health Care.* 25(2), 181-189.

Lubitz, J., Beebe, J. and Baker, C. (1995) Longevity and Medicare expenditures. *New England Journal of Medicine.* 332(15), 999-1003.

Maynard, A. (1991) Developing the health care market. *Economic Journal.* 101(408), 1277-1286.

Meltzer, D. (1997) Accounting for future costs in medical cost-effectiveness analysis. J*ournal of Health Economics.* 16(1), 33-64.

NICE (National Institute for Health and Care Excellence). (2013a) Guidelines to the methods of technology appraisals 2013. Available at http://publications.nice.org.uk/guide-to-the-methods-of-technology-appraisal-2013-pmg9. [Accessed 9 August 2013].

NICE (National Institute for Health and Care Excellence). (2013b) Technology appraisals topic selection. Available at http://www.nice.org.uk/getinvolved/topicselection/TechnologyAppraisalTopicSelection.jsp. [Accessed 9 August 2013].

NICE (National Institute for Health and Care Excellence). (2013c) What we do. Available at http://www.nice.org.uk/aboutnice/whatwedo/what_we_do.jsp.[Accessed 9 August 2013].

O'Brien, B.J., Drummond, M.F., Labelle, R.J. and Willan, A. (1994) In search of power and significance: Issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Medical Care.* 32(2), 150-163.

OECD. (2013) Health: Key tables from OECD. Available at doi 10.1787/hlthxp-total-table-2012-2-en. [Accessed 24 July 2013].

PBAC (Pharmaceutical Benefits Advisory Committee). (2008) *Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee.* Version 4.3. Canberra: Commonwealth of Australia. Available at http://www.pbs.gov.au/info/industry/listing/elements/pbac-guidelines [Accessed 9 August 2013].

Sudlow, C.L. and Counsell, C.E. (2003) Problems with UK government's risk sharing scheme for assessing drugs for multiple sclerosis. *British Medical Journal.* 326(7385), 388-392.

Sullivan, S.D., Lyles, A., Luce, B. and Grigar, J. (2001) AMCP guidance for submission of clinical and economic evaluation data to support formulary listing in US health plans and pharmacy benefits management organizations. *Journal of Managed Care Pharmacy.* 7(4), 272-282.

Swedberg, K., Kjekshus, J. and Snapinn, S. (1999) Long-term survival in severe heart failure in patients treated with enalapril. Ten year follow-up of CONSENSUS I. *European Heart Journal.* 20(2), 136-139.

Task Force (Task Force on Principles for Economic Analysis of Health Care Technology). (1995) Economic analysis of health technology assessment: A report on principles. *Annals of Internal Medicine.* 123(1), 61-70.

Upmeier, H. and Miltenburger, C. (2000) *The cost of multiple sclerosis in Germany: A top down analysis.* Poster presentation at the 16th Annual Meeting of the International Society of Technology Assessment in Health Care. The Hague. 18-21 June.

Van Hout, B.A., Al, M.J., Gordon, G.S. and Rutten, F.F. (1994) Costs, effects and C/E-ratios alongside a clinical trial. *Health Economics.* 3(5), 309-319.

Verstappen, S.M., Boonen, A., Verkleij, H., Bijlsma, J.W., Buskens, E. and Jacobs, J.W. (2005) Productivity costs among patients with rheumatoid arthritis: The influence of methods and sources to value loss of productivity. *Annals of the Rheumatic Diseases.* 64(12), 1754-1760.

Wakker, P. and Klaassen, M.P. (1995) Confidence intervals for cost/effectiveness ratios. *Health Economics.* 4(5), 373-381.

WHO (World Health Organisation). (2011) *World health statistics.* Geneva: WHO.

WHO (World Health Organization). (1948) Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19–22 June 1946 and entered into force on 7 April 1948. Geneva: World Health Organization.

Wilking, U., Jönsson, B., Wilking, N. and Bergh, J. (2010) Trastuzumab use in breast cancer patients in the six Health Care Regions in Sweden. *Acta Oncolofica.* 49(6), 844-850.

Zethraeus, N., Johannesson, M. and Jönsson, B. (1999) A computer model to analyze the cost-effectiveness of hormone replacement therapy. *International Journal of Technology Assessment in Health Care.* 15(2), 352-365.

FURTHER READING

General Health Economics and the Economic Evaluation of Health Care

Drummond, M.F., Sculpher, M.J., Torrance, G.W., O'Brien, B.J. and Stoddart, G.L. (2005) *Methods for the economic evaluation of health care programmes.* Third edition. Oxford: Oxford Medical Publications.

Folland, S., Goodman, A.C. and Stano, M. (2012) *The economics of health and health care.* Seventh Edition. Oxford: Maxwell MacMillan International.

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

Goodman, C.S. *HTA 101: Introduction to health technology assessment.* Washington, D.C.: National Information Center on Health Service Research and Health Care Technology. Available at http://www.nlm.nih.gov/nichsr/hta101/ta101_c1.html. [Accessed 13 August 2013].

Johannesson, M. (1996) *Theory and methods of economic evaluation of health care.* Dordrecht: Kluwer Academic Publishers.

Morris, S., Devlin, N.J., Parkin, D. and Spencer, A. (2012) *Economic analysis in health care.* Second edition. Chichester: John Wiley and Sons, Ltd.

Phelps, C.E. (2012) *Health economics.* Fifth Edition. New York: Harper Collins Publishers Inc.

Sloan, F.A. ed. (1995) *Valuing health care: Costs, benefits and effectiveness of pharmaceuticals and other medical technologies.* New York: Cambridge University Press.

Confidence Intervals for Cost-effectiveness Ratios

Briggs, A.H., Mooney, C.Z. and Wonderling, D.E. (1999) Constructing confidence intervals for cost-effectiveness ratios: An evaluation of parametric and non-parametric techniques using Monte Carlo simulation. *Statistics in Medicine.* 18(23), 3245-3262.

Briggs, A.H., Wonderling, D.E. and Mooney. C.Z. (1997) Pulling cost-effectiveness analysis up by its bootstraps: A non-parametric approach to confidence interval estimation. *Health Economics.* 6(4), 327 -340.

Polsky, D., Glick, H.A., Willke, R. and Schulman, K. (1997) Confidence intervals for cost-effectiveness ratios: A comparison of four methods. *Health Economics.* 6(4), 243-252.

Tambour, M. and Zethraeus, N. (1998) Bootstrap confidence intervals for cost-effectiveness ratios: Some simulation results. *Health Economics.* 7(2), 143–147.

Wakker, P. and Klaassen, M.P. (1995) Confidence intervals for cost-effectiveness ratios. *Health Economics.* 4(5), 373–381.

Cost-effectiveness Acceptability Curves

Barton, G., Briggs, A.H. and Fenwick, E. (2008) Optimal cost-effectiveness decisions: The role of the cost-effectiveness acceptability curve (CEAC), the cost-effectiveness acceptability frontier (CEAF), and the expected value of perfect information (EVPI). *Value in Health.* 11(5), 886–897.

Briggs, A.H. and Fenn, P. (1998) Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. *Health Economics.* 7(8), 723–740.

Fenwick, E., Claxton, K. and Sculpher M. (2001) Representing uncertainty: The role of cost-effectiveness acceptability curves. *Health Economics.* 10(8), 779-787.

Fenwick, E., O'Brien, B.J. and Briggs, A.H. (2004) Cost-effectiveness acceptability curves—facts, fallacies and frequently asked questions. *Health Economics.* 13(5), 405–415.

Lothgren, M. and Zethraeus, N. (2000) Definition, interpretation and calculation of cost-effectiveness acceptability curves. *Health Economics.* 9(7), 623-630.

Van Hout, B.A., Malwenn, J.A., Gordon, G.S. and Rutten, F.F.H. (1994) Costs, effects and C-E ratios alongside a clinical trial. *Health Economics.* 3(5), 309–319.

Costs

Briggs, A.H. and Gray, A. (1998) The distribution of health care costs and their statistical analysis for economic evaluation. *Journal of Health Services Research and Policy.* 3(4), 233-245.

Desgagne, A., Castilloux, A-M., Angers, J-F. and LeLorier, J. (1998) The use of the bootstrap statistical method for the pharmacoeconomic cost analysis of skewed data. *PharmacoEconomics.* 13(5), 487–497.

Luce, B.R. and Elixhauser, A. (1990) Estimating costs in economic evaluation of medical technologies. *International Journal of Technology Assessment in Health Care.* 6(1), 57-75.

Decision Analysis

Briggs, A.H., Sculpher, M. and Claxton, K. (2006) D*ecision modelling for health economic evaluation.* Oxford: Oxford University Press.

Keeney, R. and Raiffa, H. (1993) *Decisions with multiple objectives: Preferences and value trade-offs.* Cambridge: Cambridge University Press.

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

Sculpher, M., Fenwick, E. and, Claxton K. (2000) Assessing quality in decision analytic cost-effectiveness models: A suggested framework and example of application. *PharmacoEconomics.* 17(5), 461–477.

Stahl, J.E. (2008) Modelling methods for pharmacoeconomics and health technology assessment. *PharmacoEconomics.* 26(2), 131-148.

Tom, E. and Schulman, K.A. (1997) Mathematical models in decision analysis. *Infection Control and Hospital Epidemiology.* 8(1), 65-73.

Weinstein, M.C. (1990) Principles of cost-effective resource allocation in health care organisations. *International Journal of Technology Assessment in Health Care.* 6(1), 93-103.

Discounting

Brouwer, W., van Hout, B. and Rutten, F.F.H. (2000) A fair approach to discounting future effects: Taking a societal perspective. *Journal of Health Services Research and Policy.* 5(2), 114-118.

Coyle, D. and Tolley, K. (1992) Discounting of health benefits in the pharmacoeconomic analysis of drug therapies. An issue for debate? *PharmacoEconomics.* 2(2), 153-162.

Krahn, M. and Gafni, A. (1993) Discounting in the economic evaluation of health care interventions. *Medical Care.* 31(5), 403-418.

Smith, D.H. and Gravelle, H. (2001) The practice of discounting in economic evaluations of health care interventions. *International Journal of Technology Assessment in Health Care.* 17(2), 236-243.

van Hout, B.A. (1998) Discounting costs and effects: A reconsideration. *Health Economics.* 7(7), 581-594.

Economic Evaluation in Clinical Practice

Eisenberg, J. M. (1989) Clinical economics: A guide to the economic analysis of clinical practices. *Journal of the American Medical Association.* 262(20), 2879-2886.

Weinstein, M.C. (1981) Economic assessment of medical practices and technologies. *Medical Decision Making.* 1(4), 309-330.

Guidelines for Economic Evaluation

Academy of Managed Care Pharmacy (AMCP). (2009) *A format for submission of clinical and economic evidence of pharmaceuticals in support of formulary consideration.* Alexandria, VA: AMCP.

Alban, A., Gyldmark, M., Pedersen, A.V. and Søgaard, J. (1997) The Danish approach to standards for economic evaluation methodologies. *PharmacoEconomics.* 12(6), 627-636.

Alban, A., Keiding, H. and Søgaard, J. (1998) *Guidelines for economic evaluations of pharmaceuticals.* Copenhagen: Danish Ministry of Health.

Canadian Agency for Drugs and Technologies in Health (CADTH). (2006) *Guidelines for the economic evaluation of health technologies: Canada.* Ottawa: CADTH.

Capri, S., Ceci, A., Terranova, L., Merlo, F. and Mantovani, L. (2001) Guidelines for economic evaluations in Italy: Recommendations from the Italian group on pharmacoeconomic studies. *Drug Information Journal.* 35(1), 189-201.

Collège des Economistes de la Santé. (2004) *French guidelines for the economic evaluation of health care technologies.* Paris: Collège des Economistes de la Santé.

College voor zorgverzekeringen (CVZ). (2006) *Guidelines for pharmacoeconomic research, updated version.* Diemen: CVZ.

Da Silva, E.A., Pinto, C.G., Sampaio, C., Pereira, J.A., Drummond, M. and Trindade, R. (1998) *Guidelines for economic drug evaluation studies.* Lisbon: Instituto Nacional da Farmácia e do Medicamento.

Dental and Pharmaceutical Benefits Agency, Sweden (TLV). (2003) *General guidelines for economic evaluations from the Pharmaceutical Benefits Board.* Stockholm: TLV.

Institute for Quality and Efficiency in Health Care (IQWiG). (2009) *General methods for the assessment of the relation of benefits to costs.* Cologne: IQWiG.

Lopez-Bastida, J., Oliva, J., Antonanzas, F., Garcia-Altes, A., Gisbert, R., Mar, J. and Puig-Junoy, J. (2010) Spanish recommendations on economic evaluation of health technologies. *European Journal of Health Economics.* 11(5), 513-520.

Ministry of Social Affairs and Health, Finland. (2009) *Decree of the Ministry of Social Affairs and Health on applications and price notifications made to the Pharmaceuticals Pricing Board—Appendix: Guidelines for preparing a health economic evaluation.* Helsinki: Ministry of Social Affairs and Health.

National Institute for Health and Care Excellence (NICE). (2013) *Guide to the methods of technology appraisal 2013.* London: NICE

Norwegian Medicines Agency. (2012) *Guidelines on how to conduct pharmacoeconomic analyses.* Oslo: Norwegian Medicines Agency.

Pharmaceutical Benefits Advisory Committee (PBAC). (2008) *Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee.* Version 4.3. Canberra: Commonwealth of Australia.

Pharmaceutical Management Agency (PHARMAC). (2007) *Prescription for pharmacoeconomic analysis: Methods for cost-utility analysis.* Wellington: PHARMAC.

Markov Analysis

Briggs, A.H. and Sculpher, M. (1998) An introduction to Markov modelling for economic evaluation. *PharmacoEconomics.*13(4), 397-409.

Hunink, M.G.M., Bult, J., DeVries, J. and Weinstein, M.C. (1997) Uncertainty in cost-effectiveness analysis evaluated with a non-parametric bootstrap method: Combining first and second-order Monte-Carlo simulation of a Markov model. *Medical Decision Making.* 17(4), 517-544.

Sonnenberg, F.A. and Beck, J.R. (1992) Markov models in medical decision making: A practical guide. *Medical Decision Making.* 13(4), 322-328.

Meta-analysis

Cochrane Collaboration. (2002) Open learning module 12. Available at http://www.cochrane-net.org/openlearning/html/mod12.htm. [Accessed 13 August 2013].

Egger, M., Smith, G.D. and Phillips, A.N. (1997) Meta-analysis: Principles and procedures. *British Medical Journal.* 315(7121), 1533-1537.

Mills, E.J., Thorlund, K. and Ioannidis, J.P.A. (2013) Demystifying trial networks and network meta-analysis. *British Medical Journal.* 346, f2914.

Net Benefits

Briggs, A.H. (2001) A Bayesian approach to stochastic cost-effectiveness analysis. *International Journal of Technology Assessment in Health Care.* 17(1), 69-82.

Laska, E.M., Meisner, M., Siegel, C. and Stinnett, A.A. (1999) Ratio-based and net benefit-based approaches to health care resource allocation: Proofs of optimality and equivalence. *Health Economics.* 8(2), 171-174.

Stinnett, A.A. and Mullahy, J.A. (1998) Net health benefits: A new framework for the analysis of uncertainty in cost-effectiveness analysis. *Medical Decision Making.* 18(2suppl), S68-S80.

QALY League Tables

Chapman, R.H., Stone, P.W., Sandberg, E.A., Bell, C. and Neumann, P.J. (2000) A comprehensive league table of cost-utility ratios and sub-table of "panel worthy" studies. *Medical Decision Making.* 20(4), 451-467.

Mason, J.M. (1994) Cost-per-QALY league tables: Their roles in pharmacoeconomic analysis. *PharmacoEconomics.* 5(6), 472-481.

Mason, J., Drummond, M. and Torrance, G. (1993) Some guidelines on the use of cost-effectiveness league tables. *British Medical Journal.* 306(6877), 570-572.

Petrou, S., Malek, M. and Davey, P.G. (1993) The reliability of cost-utility estimates in cost-per-QALY league tables. *PharmacoEconomics.* 3(5), 345-353.

Quality of Life

Brazier, J., Ratcliffe, J., Salomon, J. and Tsuchiya, A. (2007) *Measuring and valuing health benefits for economic evaluation.* Oxford: Oxford University Press.

Gold, M.R., Stevenson, D. and Fryback, D.G. (2002) HALYs and QALYs and DALYs, Oh My: Similarities and differences in summary measures of population health. *Annual Reviews Public Health.* 23, 115-134.

Guyatt, G.W., Feeny, J.H. and Patrick, D.L. (1993) Measuring health-related quality of life. *Annals of Internal Medicine.* 118(8), 622-629.

Spilker, B. ed. (1996) *Quality of life and pharmacoeconomics in clinical trials.* Second edition. Philadelphia: Lippincott-Raven.

Walker, S.R. and Rosser, R.M. eds. (1993) *Quality of life assessment: Key issues in the 1990s.* Dordrecht: Kluwer Academic Publishers.

Sensitivity Analysis in Economic Evaluation

Briggs, A.H. (1995) Handling uncertainty in the results of economic evaluation. *Briefing.* No. 32. London: Office of Health Economics.

Briggs, A.H. (2000) Handling uncertainty in cost-effectiveness models. *PharmacoEconomics.* 17(5), 479-500.

Briggs, A.H. and Sculpher, M. (1995) Sensitivity analysis in economic evaluation: A review of published studies. *Health Economics.* 4(5), 359-371.

Briggs, A.H., Sculpher, M. and Buxton, M.J. (1994) Uncertainty in the economic evaluation of health care technologies: The role of sensitivity analysis. *Health Economics.* 3(2), 95-104.

Statistics in Economic Evaluation

Briggs, A.H. and Gray, A.M. (1998) Power and sample size calculations for stochastic cost-effectiveness analysis. *Medical Decision Making.* 18(Suppl), S81-S92.

Coyle D. (1996) Statistical analysis in pharmacoeconomic studies: A review of current issues and standards. *PharmacoEconomics.* 9(6), 506-516.

O'Brien, B.J., Drummond, M.F., Labelle, R.J. and Willan, A. (1994) In search of power and significance. Issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Medical Care.* 32(2), 150-163.

Utility Assessment

Torrance, G.W. (1986) Measurement of health state utilities for economic appraisal: A review. *Journal of Health Economics.* 5(1), 1-30.

Williams, A. (1995) *The measurement and valuation of health: A chronicle.* Discussion paper 136. York: University of York Centre for Health Economics.

Willingness to Pay

Bala, M.V., Mauskopf, J.A. and Wood, L.L. (1999) Willingness to pay as a measure of health benefits. *PharmacoEconomics.* 15(1), 9-18.

Johannesson, M. and Jönsson, B. (1991) Economic evaluation in health care: Is there a role for cost-benefit analysis? *Health Policy.* 17(1), 1-23.

Olsen, J.A. and Smith, R.D. (2001) Theory versus practice: A review of 'willingness-to-pay' in health and health care. *Health Economics.* 10(1), 39-52.