Equity and the economic evaluation of healthcare

F Sassi
L Archard
J Le Grand
How to obtain copies of this and other HTA Programme reports.

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

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

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

You can order HTA monographs from our Despatch Agents:

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

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

Contact details are as follows:

HTA Despatch

Email: orders@hta.ac.uk
c/o Direct Mail Works Ltd
Tel: 02392 492 000
4 Oakwood Business Centre
Fax: 02392 478 555
Downley, HAVANT PO9 2NP, UK
Fax from outside the UK: +44 2392 478 555

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

Payment methods

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

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

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

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

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

F Sassi*
L Archard
J Le Grand

Department of Social Policy and LSE Health and Social Care, The London School of Economics and Political Science, London, UK

* Corresponding author

Competing interests: none declared

Published March 2001

This report should be referenced as follows:


*Health Technology Assessment* is indexed in *Index Medicus/MEDLINE and Excerpta Medical/EMBASE*. Copies of the Executive Summaries are available from the NCCHTA website (see opposite).
The NHS R&D Health Technology Assessment (HTA) Programme was set up in 1993 to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies is produced in the most efficient way for those who use, manage and provide care in the NHS.

Initially, six HTA panels (pharmaceuticals, acute sector, primary and community care, diagnostics and imaging, population screening, methodology) helped to set the research priorities for the HTA Programme. However, during the past few years there have been a number of changes in and around NHS R&D, such as the establishment of the National Institute for Clinical Excellence (NICE) and the creation of three new research programmes: Service Delivery and Organisation (SDO); New and Emerging Applications of Technology (NEAT); and the Methodology Programme.

Although the National Coordinating Centre for Health Technology Assessment (NCCHTA) commissions research on behalf of the Methodology Programme, it is the Methodology Group that now considers and advises the Methodology Programme Director on the best research projects to pursue.

The research reported in this monograph was funded as project number 96/24/03.

The views expressed in this publication are those of the authors and not necessarily those of the Methodology Programme, HTA Programme or the Department of Health. The editors wish to emphasise that funding and publication of this research by the NHS should not be taken as implicit support for any recommendations made by the authors.

**Criteria for inclusion in the HTA monograph series**

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

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

Methodology Programme Director: Professor Richard Lilford
HTA Programme Director: Professor Kent Woods
Series Editors: Professor Andrew Stevens, Dr Ken Stein, Professor John Gabbay and Dr Ruairidh Milne

Monograph Editorial Manager: Melanie Corris

The editors and publisher have tried to ensure the accuracy of this report but do not accept liability for damages or losses arising from material published in this report. They would like to thank the referees for their constructive comments on the draft document.
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of abbreviations</td>
<td>i</td>
</tr>
<tr>
<td>Executive summary</td>
<td>iii</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2 Equity and the methodology of healthcare economic evaluation</td>
<td>3</td>
</tr>
<tr>
<td>Research questions and literature review methods</td>
<td>3</td>
</tr>
<tr>
<td>The theoretical foundations of healthcare economic evaluation</td>
<td>4</td>
</tr>
<tr>
<td>Distributional implications of current methods for economic evaluation</td>
<td>14</td>
</tr>
<tr>
<td>The equity-efficiency trade-off in health and healthcare</td>
<td>16</td>
</tr>
<tr>
<td>Equity weights and economic evaluation</td>
<td>19</td>
</tr>
<tr>
<td>Equity across social groups with different levels of health</td>
<td>21</td>
</tr>
<tr>
<td>Equity across age groups</td>
<td>23</td>
</tr>
<tr>
<td>Alternative methods for incorporating equity concerns in economic evaluation</td>
<td>28</td>
</tr>
<tr>
<td>Discussion</td>
<td>30</td>
</tr>
<tr>
<td>Conclusions</td>
<td>33</td>
</tr>
<tr>
<td>3 A review of published economic evaluations</td>
<td>35</td>
</tr>
<tr>
<td>Research questions</td>
<td>35</td>
</tr>
<tr>
<td>Systematic review methods</td>
<td>35</td>
</tr>
<tr>
<td>Results of the systematic review</td>
<td>42</td>
</tr>
<tr>
<td>Discussion</td>
<td>46</td>
</tr>
<tr>
<td>4 Distributional implications of recent health policies in the UK</td>
<td>47</td>
</tr>
<tr>
<td>Introduction</td>
<td>47</td>
</tr>
<tr>
<td>The national cervical cancer screening programme in England</td>
<td>47</td>
</tr>
<tr>
<td>The allocation of kidneys for renal transplantation in the UK</td>
<td>54</td>
</tr>
<tr>
<td>Screening for haemoglobinopathies in the UK</td>
<td>61</td>
</tr>
<tr>
<td>5 Conclusions and recommendations for further research</td>
<td>67</td>
</tr>
<tr>
<td>Individual versus societal perspective in the assessment of distributional consequences</td>
<td>67</td>
</tr>
<tr>
<td>Consistency between interdependent utilities and outcome measures in cost-effectiveness analysis</td>
<td>68</td>
</tr>
<tr>
<td>Empirical issues in the assessment of interdependent utilities and social-welfare functions</td>
<td>68</td>
</tr>
<tr>
<td>Consistency between the maximands of economic analysis and policy-making</td>
<td>68</td>
</tr>
<tr>
<td>Recommendations for further research</td>
<td>69</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>71</td>
</tr>
<tr>
<td>References</td>
<td>73</td>
</tr>
<tr>
<td>Appendix 1 Electronic databases used</td>
<td>81</td>
</tr>
<tr>
<td>Appendix 2 Inclusion criteria for the review of economic evaluations</td>
<td>83</td>
</tr>
<tr>
<td>Appendix 3 Checklist for the review of economic evaluations</td>
<td>85</td>
</tr>
<tr>
<td>Appendix 4 Economic evaluations reviewed in this report</td>
<td>89</td>
</tr>
<tr>
<td>Appendix 5 Selected characteristics of the economic evaluations reviewed</td>
<td>105</td>
</tr>
<tr>
<td>Appendix 6 Cervical screening case study: sources of data and sensitivity analysis</td>
<td>129</td>
</tr>
<tr>
<td>Health Technology Assessment reports published to date</td>
<td>131</td>
</tr>
<tr>
<td>Methodology Group</td>
<td>137</td>
</tr>
<tr>
<td>HTA Commissioning Board</td>
<td>138</td>
</tr>
</tbody>
</table>
## List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAPD</td>
<td>continuous ambulatory peritoneal dialysis</td>
</tr>
<tr>
<td>CRD</td>
<td>Centre for Reviews and Dissemination</td>
</tr>
<tr>
<td>DALY</td>
<td>disability-adjusted life-year</td>
</tr>
<tr>
<td>ESRD</td>
<td>end-stage renal disease</td>
</tr>
<tr>
<td>HLA</td>
<td>human leucocyte antigen</td>
</tr>
<tr>
<td>HRSWF</td>
<td>health-related social-welfare function</td>
</tr>
<tr>
<td>HYE</td>
<td>healthy-years equivalent</td>
</tr>
<tr>
<td>IBSS</td>
<td>International Bibliography of the Social Sciences</td>
</tr>
<tr>
<td>PAIS</td>
<td>Public Affairs Information System</td>
</tr>
<tr>
<td>PTO</td>
<td>person trade-off technique</td>
</tr>
<tr>
<td>QALY</td>
<td>quality-adjusted life-year</td>
</tr>
<tr>
<td>SSCI</td>
<td>Social Science Citation Index</td>
</tr>
<tr>
<td>UKTSSA</td>
<td>UK Transplant Support Service Authority</td>
</tr>
</tbody>
</table>
Background

Cost–benefit and cost-effectiveness analysis techniques have been applied extensively to healthcare resource allocation problems since the 1960s. Hundreds of economic evaluations are currently published every year in major medical and health services research journals and are easily available to potential users. However, the role played by such evaluations in healthcare decisions is still very limited due to a number of supply-side and demand-side factors, among which the nature and the assumptions of the methods used are of primary importance. In particular, much attention has been placed in recent years on the limited ability of cost–benefit and cost-effectiveness analysis to reflect social values. Despite having been developed as normative tools, economic evaluation techniques tend to guide decision-makers towards the maximisation of health gains within a resource constraint, regardless of which individuals or population groups may benefit from a health intervention or perhaps be penalised by that intervention. Distributional effects seem to have been completely neglected in existing economic evaluations, thus ignoring the equity dimension of resource-allocation problems.

Aims

The aims of this project were threefold:

- to review the methodological solutions proposed for addressing equity concerns through economic evaluation and to determine whether these are consistent with the theoretical foundations of economic evaluation, whether they are practically viable, and whether their adoption would be sufficient to confer normative strength to the results of economic analyses
- to assess whether and how the potential distributional effects of resource allocation decisions have been taken into consideration in existing economic evaluations
- to examine the cost-effectiveness and the distributional implications of selected healthcare policies currently in use in the UK, with the aim of identifying possible equity–efficiency trade-offs and determining how these have been dealt with in the absence of appropriate analyses.

Methods

First, a systematic review of the methodological literature and of the evidence about individual and collective preferences towards different equity dimensions in health and healthcare was conducted. The solutions proposed for incorporating an equity dimension into economic evaluation were described and assessed in the light of the theoretical foundations of economic evaluation and of revealed preferences for alternative distributions.

A systematic review of empirical economic evaluations published in five sample years (1987, 1992, 1995, 1996, 1997) was then conducted using electronic search strategies specifically developed and tested for the systematic retrieval of economic evaluations from the main literature databases. Studies were examined by means of a checklist devised for the purpose of identifying whether these had answered key distributional questions.

Finally, three case studies of healthcare policies adopted in the UK were conducted using a combination of literature review and primary research methods. The three policies were: cervical cancer screening, the central allocation of kidneys for renal transplantation, and neonatal screening for sickle cell disease. The cost-effectiveness of each of the three policies was examined alongside distributional implications, regarding socio-economic status in the first case study, age in the second, and ethnicity in the third.

Results

The methodological solution for addressing distributional concerns in economic evaluation that has attracted the interest of health economists more than any other involves weighting health outcomes for specific equity dimensions (e.g. age, socio-economic condition). Other solutions that were explored include the use of willingness-to-pay measures, the person trade-off technique (all belonging to the normative category – following a general distinction between ‘normative’ or ‘positive’ made by Mishan about allocation economics – as well as equity weighting) and a positive solution based on the tabulation of the effects of health interventions in different subpopulations.
The conclusion of this systematic review of the methodological literature is that the normative route to addressing equity concerns through economic evaluation poses significant, if not insurmountable, theoretical and practical problems. Normative solutions can be based on the measurement of interdependent utilities, on the definition of a social welfare function, or both. The former alone is not consistent with cost-effectiveness techniques based on measures of health gain that do not necessarily reflect utilities for own consumption of healthcare and would lead to allocations that are arguably undesirable; whereas defining a social welfare function would involve extremely complex measurements that are far beyond the reach of existing studies of individual and collective values. A positive solution to addressing equity concerns seems a more appropriate way forward for economic evaluation. This would entail presenting essential information on the effects of health interventions in different population groups to decision-makers who would ultimately apply their own values and trade-offs and make decisions accordingly.

The picture resulting from the review of empirical studies was extremely disappointing. None of the economic evaluations examined in the review provided enough information to allow decision-makers to judge the distributional consequences of alternative resource allocations. The studies examined were unsuitable for assessing any of three key distributional effects:

- the effects of switching between the (mutually exclusive) interventions compared in an evaluation
- the effects of providing an intervention selectively to a subset of the overall population that may potentially benefit from that intervention
- the effects of prioritising between interventions competing for a given pool of resources.

Only half of the economic evaluations reviewed measure outcomes that can be meaningfully used in comparisons of cost-effectiveness ratios across interventions. Only about one in eight of these studies report some information on the characteristics of the population that may benefit from the interventions appraised. A larger number of studies report information on the effects and cost-effectiveness of interventions in specific patient subgroups defined in terms of age, gender, risk profile or ethnic group. This information may help in assessing distributional effects of the second type mentioned above. However, these cost-effectiveness ratios should be calculated through direct comparisons between patient groups rather than indirectly, through comparisons of alternative interventions in different groups, as in all the evaluations reviewed.

The three case studies show that a different emphasis on the equity dimension has been placed in different policy choices, and the equity principles that seem to have guided such policies vary significantly. The policies examined do not always reflect the social values elicited by the empirical studies reviewed in the first part of this report. This appears to be the case even when the process through which a policy is developed is more explicit and the evidence base is relatively strong (as in the renal transplantation case study), although it is probably fair to say that a wider availability of information on the likely distributional consequences of alternative policy options would, in many cases, have led to different policy choices.

Conclusions

The main conclusion of this project is that existing economic evaluations do not represent an adequate guide to resource allocation decisions when the distributional effects of such decisions may be relevant. Not only do they not address explicitly the equity dimension, but they do not even provide the information that decision-makers would need to make a judgement on the desirability of alternative allocations.

Recommendations for future research

The normative route advocated by many as a means of addressing equity concerns through the development of existing economic evaluation techniques does not appear to be a viable solution at present. The evidence base about social values for alternative distributions is largely insufficient, and empirical research methods aimed at determining suitable social welfare functions need to be strengthened (particularly with regard to framing effects and multi-dimensional measurements). The research required to close these gaps will inevitably take several years to complete, and economic evaluation cannot wait for the uncertain outcomes of such research. Methodological developments aimed at incorporating an equity dimension into economic evaluations are required as a matter of priority. Therefore, a short-term solution is proposed, based on existing approaches not widely applied in practice. These involve the systematic gathering of information about the expected distributional effects of resource allocation.
decisions (e.g. characteristics of the populations that may benefit from the health interventions appraised, information on the effectiveness and cost-effectiveness of the interventions in different subgroups). This information should be presented to decision agents who will weight such effects on the basis of their own objective functions. It would seem reasonable to pursue this solution for the immediate development of economic evaluation, while more theoretical, methodological and empirical research is undertaken to determine the viability of a normative approach.
Economic analysis techniques have been applied in the field of healthcare for just over three decades with the aim of improving the efficiency of resource allocation. Although the basic concept of economic evaluation has not fundamentally changed over time, significant methodological developments have taken place and new forms of analysis, almost specific to healthcare, have progressively emerged. Such developments are now often regarded as a model in the application of economic analysis to areas of public intervention in which efficiency had previously been less of a concern. However, there are still important methodological gaps and controversies that limit the potential and credibility of economic evaluation as a decision support tool. These concur with other factors (e.g. inappropriate timing of evaluations, poor communication of results) to increase the scepticism of decision-makers towards economic analysis and to prevent them from using economic evidence in their decisions.

One of the main methodological deficiencies that places economic evaluations at odds with the values and principles informing healthcare decisions is the neglect of the equity dimension (i.e. the lack of consideration of distributional effects that may arise from resource allocation decisions). Or, even worse, the lack of an explicit indication of what distributional effects a decision based on the results of an economic evaluation would have.

Equity is increasingly a key concern of healthcare decision-makers, particularly at the policy level, and society is not indifferent to alternative distributions of health and healthcare among individuals and population groups. By neglecting this dimension, economic analysis loses much of its normative power and restricts itself to a relatively narrow role in supporting healthcare decisions. Many health economists have recently placed their attention on possible ways of addressing equity concerns regarding health and healthcare through economic evaluation. The solution that has attracted more interest involves weighting health outcomes for specific equity dimensions (e.g. age, socio-economic condition), although this is not the only option explored.

This report reviews the methodological solutions proposed for addressing equity concerns through economic evaluation and determines whether these are consistent with the theoretical foundations of economic evaluation, whether they are practically viable, and whether their adoption would be sufficient to confer normative strength on the results of economic analyses.

The report is structured in three main parts. Chapter 2 is devoted to the analysis of the methodological literature and of the evidence about individual and collective preferences towards different equity dimensions in health and healthcare. The solutions proposed for incorporating an equity dimension in economic evaluation are described and assessed in the light of the theoretical foundations of economic evaluation and of revealed preferences for alternative distributions. Chapter 3 is devoted to the analysis of empirical economic evaluations. The results of a systematic review of economic evaluations published in five sample years are reported, illustrating whether and how the potential distributional effects of resource allocation decisions have been taken into consideration. Finally, chapter 4 includes three case studies of healthcare policies adopted in the UK: cervical cancer screening, renal transplantation and neonatal screening for sickle cell disease. The cost-effectiveness of each of the three policies is examined alongside distributional implications (regarding socio-economic status in the first case study, age in the second, and ethnicity in the third), and implicit values with regard to equity are inferred where possible. Chapters 2 and 3 are based on systematic reviews of the existing literature, whereas the case studies reported in chapter 4 combine literature reviews with varying degrees of original empirical analysis.
Chapter 2

Equity and the methodology of healthcare economic evaluation

Research questions and literature review methods

In this section we report the results of a broadly based review of the health economics literature aimed at answering the following key research questions:

- Can the theoretical principles on which economic evaluation is based enable distributional judgements to be incorporated in the analysis and normative statements to be made accordingly?
- What theories of justice and equity principles could economic evaluation draw upon in addressing distributional concerns?
- What distributional implications do current economic evaluation methods have when equity issues are not explicitly addressed?
- What are the preferences of individuals and society with regard to different dimensions of equity in health and healthcare?
- Can preferences for different dimensions of equity be incorporated in economic evaluation by applying appropriate weights to health outcomes?
- Can methods other than equity weights be used to address distributional concerns in economic evaluation?

Three broad areas were identified for the literature review:

- theoretical foundations of economic evaluation (including any literature on the theoretical status of equity in economic evaluation)
- the distributional implications of conventional economic evaluation (including theories of justice applicable to resource allocation in healthcare)
- empirical literature relevant to equity in economic evaluation (including surveys of decision-maker or public opinion towards distributional issues, studies to elicit equity weights, and the application of equity weights).

Relevant papers in each of these areas were identified in three ways:

- the knowledge of the research team
- database searches
- bibliographies of papers acquired through the above two.

The following databases were searched: MEDLINE, International Bibliography of the Social Sciences (IBSS), Social Science Citation Index (SSCI), PsycLIT, EconLit. Different sets of search terms were required for each database to take account of the different indexing systems and subjects served by the databases. Inclusion criteria of the sort we applied to potential economic evaluations were not appropriate to this kind of literature search. Whether an article was relevant was a matter for judgement on a case-by-case basis.

**MEDLINE search terms**
1. equity [tw]
2. priority-setting [tw]
3. fair* [tw]
4. distribut* [tw]
5. ethics [tw]
6. QALY* [tw]
7. Health maximisation [tw]

Terms 1 to 5 were combined with the ‘OR’ operator; terms 6 and 7 were combined with the ‘OR’ operator. The intersection was then found between these two groups using the ‘AND’ operator.

**IBSS/SSCI search terms**
1. Health care
2. Equity
3. Rationing
4. Justice

Terms 2 to 4 were combined with the ‘OR’ operator. We then searched for the intersection with term 1.

**EconLit search terms**
1. Equity
2. Cost-effective*
3. Cost-benefit
4. Cost-utility analysis
5. Economic evaluation
Terms 2 to 5 were combined with the ‘OR’ operator. We searched for the intersection with term 1.

PsycLIT search terms
PsycLIT was searched through the intersection between the term ‘Equity’ with the term ‘Health Care’.

The theoretical foundations of healthcare economic evaluation

The earliest healthcare economic evaluation referenced by health economics bibliographies is a study of the cost of tuberculosis published in 1920. However, Sorkin sets the earliest work as long ago as 1699, when Sir William Petty calculated “the rate of return obtained by moving the London population outside the city during epidemics of the plague”. Other early evaluative studies in the twentieth century include some on the cost of mental illness, road accidents, alcoholism, and various categories of disease.

The first study on healthcare explicitly labelled ‘cost–benefit’ is probably one on water supply investments, although many of those previously cited did calculate what are, in practice, cost–benefit ratios of interventions. An early attempt to compare different strategies for a specific condition is the study on syphilis control programmes by Klarman. All these studies were cost–benefit analyses and adopted the human capital approach to measuring benefits, according to which resources devoted to healthcare can be seen as an investment in health, making more labour available to the economy. The conceptual limitations of this approach soon became evident to many authors. An alternative approach to assessing health outcomes in cost–benefit analysis, based on the measurement of the willingness to pay of individuals for health gains showed poor viability on empirical grounds due to difficulties and biases involved in attaching monetary values to human lives and health. The apparently simpler framework of cost-effectiveness analysis was preferred by many researchers. The first two studies of this type, concerning treatment of chronic renal disease and maternal and child care, were published in 1968. From 1966 to 1973, on average, 42.1% of economic evaluations were cost-effectiveness analyses. That figure progressively increased to 53.2% in the period 1974–78, and to 59% and 64% in the periods 1979–84 and 1985–90, respectively. The actual proportion of cost-effectiveness analyses is probably even larger than this, as the reviews cited are not based on a detailed examination of the relevant studies, and cost–benefit labels often hide cost-effectiveness studies.

A major methodological breakthrough was achieved in the early 1970s by linking Von Neumann–Morgenstern expected utility theory to economic evaluation. This gave rise to cost–utility analysis, which is considered by many authors as a form of cost-effectiveness analysis. The first empirical work based on utility measures is one on phenylketonuria screening, in which a rating scale technique was employed, although subjective quality-adjustment of life-years had been used before by Klarman. Conceptual and empirical difficulties in cost–utility analysis prevented a rapid proliferation of this approach; an extensive review published in 1992 counted 51 studies of this type, most of which were methodologically deficient. A recent update of this review showed increasing numbers of studies, but continuing problems in their methodological quality. However, the cost–utility/cost-effectiveness framework is now widely accepted as the reference standard in healthcare economic evaluation, although some authors deem its limitations overwhelming and interest in cost–benefit analysis and the willingness-to-pay approach is mounting again.

Since the 1960s the basic concept of economic evaluation in healthcare has not changed, despite important methodological developments, as illustrated by the following three definitions:

- Essentially, cost–benefit analysis entails a comparison of costs and benefits for a series of programmes thought of as alternative or competitive.

- Economic evaluation is the comparative analysis of alternative courses of action in terms of both their costs and consequences.

- Cost-effectiveness analysis is a method designed to assess the comparative impacts of expenditures on different health interventions.

All three definitions have in common the emphasis on two elements: the comparative nature of economic evaluation (alternatives must be considered), and the attention placed on both costs and outcomes of such alternatives. Whether the application of this basic concept in formulating guidance to healthcare resource allocation decisions may or may not lead to desirable social outcomes has been a matter of intense debate, particularly in the 1990s.
This debate has been mainly centred on the welfare economics foundations of economic evaluation, and is briefly reported in the following sections.

**Welfare economics and cost–benefit analysis**

The theoretical foundations of economic evaluation techniques, as currently applied in healthcare, can be traced in disciplines such as economics and operational research. In particular, a link can be established with welfare economics and optimisation theory. Expected utility theory, which presents large areas of overlap with the former two, is also a foundation of some of the developments of economic evaluation.

Welfare economics is essentially normative allocation economics* and is concerned with formulating and justifying propositions by which alternative economic situations may be ranked. The starting point of welfare economics is the well-being of individuals in terms of the utility they derive from different states of the world, and the final aim is “achieving a social maximum derived from individual desires”. The general equilibrium analysis typical of welfare economics is based on a number of assumptions, among which is the postulate that individuals aim at maximising a well-defined utility function, based on preferences for the consumption of different goods, and that social welfare is a function of such individual preferences. The postulated utility-maximising attitude of individuals is described by Sen as “self-interested behaviour” and is characterised by three features:

- **Self-centred welfare:** a person’s welfare depends only on his or her own consumption ...
- **Self-welfare goals:** a person’s goal is to maximise his or her own welfare and – given uncertainty – the probability-weighted expected value of that welfare ...
- **Self-goal choice:** each act of choice of a person is guided immediately by the pursuit of one’s own goal ...

These assumptions appear untenable when dealing with health and healthcare. We shall discuss later how this may affect the validity of the welfare economics framework as a basis for economic evaluation in healthcare.

Welfare economics defines an efficient allocation of resources (Pareto optimality) as one where any reallocation would make at least one individual worse off. Therefore a Pareto improvement is the result of a reallocation of resources which makes at least one individual better off, without making anyone worse off. Pareto optimality is reached when efficiency in consumption and efficiency in production are simultaneously achieved. Analytically, the former is achieved when the marginal rate of substitution† between any two goods is the same across individuals, whereas the latter is achieved when the marginal rate of technical substitution between factors of production‡ is the same across production processes. Thus, in a simplified society with two individuals and two goods, Pareto optimality is represented by the points on a utility possibilities frontier (Figure 1) whereby the marginal rate of substitution between the two goods for the two individuals equals the marginal rate of transformation§ in the production of the two goods. All points on the utility possibilities frontier indicate an efficient allocation of societal resources.

As Reinhardt points out, however, economists whose judgement is guided by the Pareto optimality criterion are unable to identify a truly ‘optimal’ point on the utility possibilities frontier. Even more importantly, faced with a choice between a point beneath the frontier and one on

---

* The distinction between normative and positive solutions, widely adopted in this report, reflects the use that Mishan makes of these terms. Mishan himself poses the question of whether a truly positive allocation economics may exist. In fact, we point out several times in this report that economic analysis is never entirely value-free. However, the distinction between normative and positive solutions to the analysis of alternative allocations is very helpful in distinguishing between solutions in which the analyst does and does not introduce his own norms in assessing the value of effects occurring to different population groups and in combining such effects.

† The number of units of one good an individual would be willing to trade in exchange of one unit of another good.

‡ The number of units of one factor needed to replace one unit of another factor in order to maintain a constant output.

§ The number of units of one good that can be produced with the resources freed reducing the production of another good by one unit.
the frontier, they would not necessarily see a move from the former to the latter as an improvement. Economists accepting the Pareto optimality criterion have very limited room for manoeuvre in determining what is desirable from a societal viewpoint. This is the price they have to pay for an apparently value-free framework that appears as the last resort to preserve the normative nature of welfare economics in increasingly ‘pluralistic’ western societies. As Sudgen and Williams indicate: “In welfare economics, social welfare has (at least) two dimensions: economic efficiency and distributional justice. The former concerns the size of the total wealth of all members of a community, while the latter concerns the way the total wealth is shared amongst individuals”. Many economists (e.g. Kaldor, as cited by Mishan) have emphasised the need to keep these two dimensions separate and to let economics address the former, whereas any judgements on the distributional implications of alternative allocations should be left to the political process.

The idea of abstracting from distributional implications in economic analysis holds when the broader potential Pareto improvement criterion (or Kaldor–Hicks principle) is adopted instead of the classical Pareto improvement criterion. According to this, a reallocation of resources which makes someone better off and someone worse off represents an improvement as long as the value of the gains exceeds the value of the losses. In other words, there is an improvement as long as the losers could be compensated by the gainers if a transfer were possible at no cost. This principle has been much criticised (e.g. Reinhardt) because of its implicit distributional implications, which are made worse by the principle that compensatory transfers are not required to take place in practice. However, it does provide a clearer (compared to the classical Pareto improvement criterion) indication of what allocations are efficient (those lying on the possibilities frontier) and what allocations are not. Moreover, adopting a pragmatic view of the operation of the economy, Mishan argues that the actual circumstances in which the potential Pareto improvement criterion can be applied would automatically limit the negative effects of not requiring compensatory transfers to take place.

Cost–benefit analysis has its theoretical basis in the potential Pareto improvement criterion. It assumes that utilities can be aggregated across individuals, which implies that they are cardinal and inter-

In fact, the Pareto optimality criterion is not value free, it incorporates a social-welfare function that gives extra weighting to the better-off.
personally comparable. Judgements on the distri-
butional implications of alternative allocations are,
as far as possible, left to decision-makers. Mainly
for this reason, cost–benefit analysis should be
typically regarded as a decision support tool,
providing useful information to decision-makers
but not providing decision rules. This is consistent
with Mishan’s argument in favour of a positive,
rather than normative, allocation economics.37

Economic analysis can, at least in principle,
move beyond efficiency and take into account
distributional effects. This is typically achieved by
considering a social-welfare function determined
by a map of social indifference curves, or welfare
contours, that indicate how much utility of one
individual or social-group society would be willing
to sacrifice in order to increase the utility of
another individual or social group. A society that is
indifferent to alternative distributions will have a
social-welfare function like the one represented
in Figure 2, whereby a sacrifice of utility for A is
acceptable only if this corresponds to a gain of an
equal amount of utility for B. This is consistent
with a utilitarian (or Benthamian) view of social
welfare as a simple sum of individual utilities. At
another extreme, a society may express L-shaped
social indifference curves, as in Figure 3, indicating
that no gain of utility for A could possibly compensate
even the smallest sacrifice of utility for B, in a purely
egalitarian view. This is also consistent with the
Rawlsian ‘maximin’ principle, according to which
social welfare is determined by the welfare of the
worst-off (although Rawls’s theory of justice is
concerned with distributions of primary goods, rather
than individual utility). Only by improving the latter
can the former be improved, and welfare gains for
the better-off are of no value to society. A further
possible shape for social indifference curves is that
illustrated in Figure 4, indicating that the larger the
overall amount of individual utility, the larger the
sacrifice society is willing to impose on that individual
in order to increase someone else’s utility. The
Pareto improvement criterion itself can be inter-
preted as implying a special form of social-welfare
function, in which the welfare contours have the
same shape as the utility possibilities frontier.44

Social-welfare functions allow the identification of
an optimal point among the efficient combinations
lying on the utility possibilities frontier. This will
correspond to the point in which the frontier is
tangent to the highest possible social indifference
curve (e.g. point X in Figures 2 to 4). By taking
into consideration a social-welfare function, and
by accepting the more flexible potential Pareto
improvement criterion, the normative limitations
of cost–benefit analysis based on the classical
Pareto improvement criterion can be overcome.
However, if this approach is attractive in theory, its
practical application in economic analysis certainly
poses major, if not insurmountable, difficulties.
Utility of A

Utility of B

FIGURE 3 ‘Egalitarian’ social-welfare function

Utility of B

Utility of A

FIGURE 4 Social-welfare function with decreasing marginal utility
These are mainly related to the feasibility of eliciting decision-makers’ values and measuring trade-offs, as well as to the limited number of equity dimensions that can be considered in a social-welfare function.

**Cost-effectiveness analysis: beyond welfare economics?**

As discussed in the previous section, the normative statements of welfare economics do not fully reflect societal and decision-makers’ values. Moreover, resource allocation is not the result of one single decision based on a defined social-welfare function, but derives from a complex set of interrelated decisions made by separate decision-makers who pursue different objective functions. This is illustrated by Sudgen and Williams in their juxtaposition of the Pareitian and decision-making approaches to cost–benefit analysis:

[The Pareitian] interpretation of cost–benefit analysis sets it apart from the process by which actual decisions are taken in practice at any point in space or time. The analyst works independently of the political decision-making process and brings to his work his own independent norms.

The question examined in this section is whether the cost-effectiveness analysis approach may help in overcoming some of the limitations of the classical cost–benefit approach, and what departures from welfare economics theory are implied by cost-effectiveness analysis.

Cost-effectiveness analysis adopts a more limited view of the benefits of healthcare interventions than does cost–benefit analysis. Benefits are measured in natural units (e.g. survival rates, life-expectancy) and along one dimension only. Attempts to incorporate more outcome dimensions and to value utilities for health outcomes rather than just measuring outcomes in natural units have led to the development of a variation of cost-effectiveness analysis defined as cost–utility analysis. In its most typical form, cost–utility analysis adopts a bidimensional outcome measure, quality-adjusted life-expectancy (measured in quality-adjusted life-years (QALYs)) which combines life-expectancy and health-related quality of life in a single index. Utilities are normally elicited for the latter, while life-expectancy is still measured in natural units. It has been argued that the QALY may be considered a measure of utility only under very restrictive assumptions, which in practice are never met.

The adoption of measures of outcome not reflecting utilities is viewed by Birch and Gafni as a reason for considering cost-effectiveness analysis (and possibly cost–utility analysis) suitable only for addressing technical efficiency issues in the production of health gains and not for allocating resources in the presence of a budget constraint. This view has been criticised by Johannesson and Weinstein, while Phelps and Mushlin have gone even further arguing that there is ‘near equivalence’ between cost–benefit and cost-effectiveness analyses as means of pursuing allocative efficiency. However, Johannesson points out that cost-effectiveness analysis may indicate when the adoption of a healthcare programme represents a Pareto improvement, or a potential Pareto improvement, only when all costs and benefits accruing to all individuals in a society are taken into account, and when the willingness to pay for the effectiveness unit is constant and the same for everyone.

The inconsistency of cost-effectiveness analysis with welfare economics principles make the normative limitations of this approach even greater than those of cost–benefit analysis illustrated in the previous section, as it makes this approach more alien to the value system of the potential users of economic evaluation. Birch and Gafni have argued that since cost-effectiveness analyses do not capture all dimensions of the outcomes gained and forgone by implementing a healthcare programme these may not be regarded as normative. While Johannesson and Weinstein agree with this view, they believe that cost-effectiveness analysis can be used to maximise a specified effectiveness objective given a finite amount of resources. QALY gains may generally be regarded as a sufficiently encompassing measure of effectiveness to overcome the problem indicated by Birch and Gafni, although some studies have demonstrated the importance of outcome dimensions that are not captured in QALY measurements (e.g. Berwick and Weinstein). However, the normative nature of cost-effectiveness analysis remains confined to the maximisation of a specified objective within a budget constraint.

The aspiration to provide a social welfare theoretical justification to the normative use of cost-effectiveness analysis in healthcare resource allocation (e.g. Garber and co-workers argue that: “to the extent that [cost-effectiveness analysis] is designed to be a practical tool for achieving societal goals, we believe that [the theory on which it is based] must be normative”) has led to the search for suitable foundations in the so-called ‘extra-welfarist’ approach. Culver has defined ‘extra-welfarism’ as a normative approach alternative to welfare economics, according to which “the
analysis of efficiency in systems like healthcare ... and in specific projects within systems ... may embrace whatever maximand(s) may be given by the customers of research or inferred by diligent enquiry by the analyst to be relevant". Therefore, the formulation of the objective function, or measure of health effects, is left to the discretion of the decision-maker, who acts as an agent for societal healthcare decisions. This move from the principles of welfare economics is justified on the following grounds:

The decision-maker occupies his position by virtue of a socially approved political process. He has been entrusted with the task of making choices on behalf of the general public, and this trust implies that he will formulate objectives for the society. He is accountable to the public for carrying out this task to their satisfaction.

However, Hurley pointed out that the extra-welfarist approach does not overcome a number of limitations of welfare economics, and may eventually leave many of the theoretical problems of economic evaluation unsolved. In particular, welfarist and extra-welfarist approaches have three main limitations in common:

- consequentialism (i.e. emphasis on outcomes rather than process)
- monism (i.e. unidimensional outcome space)
- neglect of equity.

In fact, the latter does not mean that economic evaluation is neutral from a distributional point of view. The implicit distributional consequences of cost–benefit analysis have been briefly discussed in the previous section, and those of cost-effectiveness analysis are discussed later in this chapter (see pages 14 and 15). As argued by Bleichrodt, under particular circumstances distributional concerns could be addressed by cost-effectiveness analysis. This report will examine what methods could be used to address specific equity questions and whether these methods may enhance the normative character of cost-effectiveness analysis by making this more consistent with the value system of decision-makers.

### Equity principles and theories of justice

Before exploring further the conceptual framework of healthcare economic evaluation, we briefly discuss a range of general theories and concepts of justice, and how they might apply to healthcare. If an equity dimension is to be incorporated in economic evaluation, the methods adopted to achieve this aim should allow different equity principles to be taken into account, depending on the values of decision-makers or society.

Several authors have reviewed equity theories and principles applicable to health and healthcare. Further discussion of the main theories can be found in these texts. What follows is a short discussion of the most prominent ideas.

#### Utilitarianism

Utilitarianism is the principle that actions or states are to be judged according the amount of human happiness (‘utility’) they produce or contain. The principle of utilitarianism has been defined as ‘maximising happiness’ or as aiming to bring ‘the greatest good to the greatest number’.

But is utility maximisation always fair? Sen remarked that “maximising the sum of individual utilities is supremely unconcerned with the interpersonal distribution of that sum”. There are concerns that utilitarianism would lead to distributions of resources and utility that we might regard as unfair. For example, utility in society might be increased most by increasing social inequalities in health: treating a wealthy person with a higher life-expectancy and capacity to benefit may produce greater utility than treating a poorer person with a lower life-expectancy and lower capacity to benefit.

Economic evaluation can be interpreted as utilitarianism applied to healthcare, the difference being that economic evaluation seeks to maximise health rather than utility. Maximisation of health gain has been suggested as compatible with equity goals. For instance, this might be true if those individuals with lowest health status also tended to be those with the greatest capacity to benefit from healthcare. Seeking to both reduce social inequality in health and improve overall population health would in this case be compatible goals. However, the advantaged groups in society (those with higher levels of income, education, social class) tend to have better health and may also have a superior capacity to benefit from healthcare. Thus, contrary to reducing inequality, the application of the principle of health maximisation may act to increase it. The distributional implications of economic evaluation are discussed in detail later in this chapter (see pages 14 and 15).

#### Rawls’s theory of justice and ‘maximin’

The work of Rawls is prominent in the philosophical discussion of social justice. Of first priority in his theory of justice is liberty: as much should be available to each citizen as is compatible with equal
liberty for all. Secondly, primary social goods, such as income, wealth, positions of responsibility, and self-respect, should be distributed in order to improve the position (in terms of primary goods) of the worst-off in society. This latter principle is known as the **maximin principle**.

How does health fit into this? Rawls himself did not include health as a primary social good. This was partly because health is determined by natural factors beyond human control and not just by social arrangements. Application of maximin to health has also been criticised on the grounds that the attempt to improve the health of the worst-off could impoverish society. Despite this, health economists have applied the maximin principle directly to health. Under maximin applied to health, the worst-off are identified in terms of their health status rather than in terms of primary goods, and the aim of health policy should be to improve the health of the least healthy.

**Egalitarianism**
What should egalitarians seek to equalise? Candidates include utility, liberty, resources or opportunity. Dworkin argued that egalitarianism should aim at equalising resources available to people and not at equalising welfare. These resources include talent, intelligence and health, as well as ‘primary goods’ such as income. People should be compensated for disabilities or lack of talent, for example, but not for having expensive tastes. In terms of healthcare, what might this mean? If health is a basic resource, then it suggests we should aim at equalising health through compensation for those who are less healthy. However, it would exclude measuring health outcome in terms of preferences for health, since this could conceivably lead to compensating those who already have better health than others (however, this result would not occur if a common set of valuations were applied to both those with better health and those whose health is worse). Compensation for the less healthy might be achieved through rationing treatment or access to healthcare in their favour.

Sen has argued we should aim at equalising people’s abilities to pursue valuable activities (e.g. to work, to engage in cultural activities such as playing music or sport) and to reach valuable states of being (e.g. to have high self-esteem, to have good health). He argues for the principle of equality of basic capabilities. In contrast to Dworkin, who argued for equalising resources, Sen’s ideas seem closer to the concept of equality of opportunity. Pereira has suggested that Sen’s ideas could be applied to healthcare. He remarks: “In short, with regard to equity, equality of capabilities implies equal access to health benefits”.

**Nozick’s entitlement theory**
According to Nozick’s entitlement theory, what matters for the fairness of a distribution is how that distribution was reached. Nozick argued that one was entitled to property, income and wealth acquired justly. But when are these goods acquired justly? Nozick’s view is that just acquisition occurs through voluntary transactions. Coercion is a violation of the natural right to liberty.

It is not clear how this theory applies to health and healthcare. On the one hand, collective provision of healthcare would seem to be unwarranted, since it relies on a level of coercion through the tax system. On the other, we may ask under what circumstances health or ill-health is justly acquired? Le Grand has suggested an answer to this: “if an individual’s ill health results from factors beyond his or her control then the situation is inequitable; if it results from factors within his or her control then it is equitable”. This principle has been called **equity as choice**, under which inequalities resulting from individual choice are not regarded as unfair. However, for equity as choice to be operational, the determinants of ill health that lie within, and outside, an individual’s control need to be clearly specified. There are questions as to the extent to which this is possible and who decides where the boundaries are.

**Some principles specific to ethical problems in healthcare**
Here we consider some specific ethical principles and arguments that have been applied to healthcare. These are:

- allocation according to need
- the ‘fair-innings’ argument
- the rule of rescue.

**Allocation according to need**
It is commonly held that healthcare should be allocated according to need. Although there are several possible definitions of need (for a discussion of these, see Culyer and Wagstaff) need is usually understood in terms of severity of illness. Health economists have defined need in terms of capacity to benefit from treatment (the definition implicit in economic evaluation). Need as severity of illness, although initially rejected by health economists, has come to be regarded as an equity
concern (see, e.g., Nord and co-workers\textsuperscript{68}). There are two common formulations of the principle of allocation according to need:

- equal treatment for equal need
- equal access for equal need.

These ideas reflect the view that healthcare is different from other goods and services, and should be distributed according to non-market principles.

The ‘fair-innings’ argument
As suggested by its name, the ‘fair-innings’ argument applies to the issue of fairness in the duration of different people’s lives (although one author has suggested it might be better applied to fairness in the quantity of quality-adjusted lifetime\textsuperscript{69}). There are numerous formulations of this argument (see Tsuchiya\textsuperscript{70} for a full discussion of these). Tsuchiya describes the ‘original’ fair-innings argument as follows:

The main point … is that we may set some amount of life-years as a ‘fair-innings’ and decide that, supposing we had to choose between saving the life of somebody above this age and somebody below this age, other things being equal, we should save the life of the younger person, who otherwise will not be able to enjoy the fair innings.\textsuperscript{70}

Some formulations involve a target life-span. In general terms, those who have already reached the target are to be given lower priority than those below the target. Other formulations do not specify a threshold number of years. For example, Williams\textsuperscript{69} proposes that the notion of a fair-innings may act as justification for a set of weights based on the prospects of achieving a fair-innings. The further an individual is from achieving the fair-innings, the larger the weight; the closer an individual is to achieving the fair-innings, the smaller the weight.

The rule of rescue
The word ‘rescue’ suggests saving someone’s life from imminent danger. It has been suggested that we will always prefer to save the life of someone in danger of dying than improve the quality of life of someone else whose life is not in danger, or even to saving future lives through disease-prevention programmes. This has been called the ‘rule of rescue’.\textsuperscript{71} However, the rule of rescue has been criticised as more a description of a human impulse than a principle of equity. Reflection may lead to a different basis for setting priorities. But, if accepted as an equity principle, it would tend to support the prioritisation of life-saving interventions, perhaps emergency services in particular.

Possible maximands for economic evaluation and distributional goals
The use of economic evaluation as a means of identifying efficient allocations of scarce resources involves maximising a given object (maximand) of relevance to society or to a particular decision-maker. Although welfare economics is clear about what maximand pertains in cost–benefit analysis (i.e. individual utility; see the quote from Sen\textsuperscript{39} on page 15), different maximands have been used in cost-effectiveness analysis and many others could potentially be used. When the analyst is able to infer a social-welfare function based on societal values regarding alternative distributions of the maximand, this social-welfare function becomes the object to be maximised. The issue of whether and how equity principles could be taken into consideration in economic evaluation depends closely on the nature of the maximand adopted.

In this section we review the characteristics of alternative maximands related to the consumption of healthcare that have been used or could potentially be used in cost–benefit and cost-effectiveness analysis, and we shall discuss the implications of each of these for the implementation of methods to incorporate distributional concerns in economic evaluation.

At its most basic level, economic evaluation may adopt some measure of effectiveness of healthcare as a maximand. This is typically the case in cost-effectiveness analysis, whereby outcomes are normally measured in natural units (e.g. life-years gained). This type of maximand does not fit into the framework of welfare economics, unless it is assumed that the willingness to pay for the effectiveness unit is constant and the same for everyone\textsuperscript{51} (see page 19). However, this assumption does not appear realistic when effectiveness is measured in natural units.\textsuperscript{45,46} More comprehensive measures of outcome, such as the QALY, have been developed with the specific aim of representing individual preferences towards health states and health gains (hence the definition of cost–utility analysis), but several authors have argued that the QALY is not a measure of utility.\textsuperscript{45–47} Alternative measures, such as the healthy-years equivalent (HYE), have been proposed to overcome some of the limitations of the QALY, but whether the former is fundamentally different from, or superior to, the latter is a matter of great controversy.\textsuperscript{52,53}
If one accepts the view expressed by Birch and Gafni that cost-effectiveness analysis measuring outcomes in non-preference-based natural units may only address technical (or production) efficiency questions, and any possibility of incorporating distributional concerns in this form of evaluation is precluded. The same applies to evaluations using QALYs, if these are not deemed to reflect individual preferences towards health outcomes. However, if QALYs or other measures of health gain are considered to be sufficiently close approximations of utilities, distributional judgements with regard to these measures can be made. Whether the application of equity principles to QALY distributions may lead to an improvement in social welfare is, of course, open to debate.

Utilities measured in terms of individual willingness to pay are certainly more consistent with the welfare economics framework than any maximand so far used in cost-effectiveness analysis. However, as previously indicated, welfare economics typically assumes that individuals adopt a self-interested behaviour, that they derive utility exclusively from own consumption and that no externalities are generated. All these assumptions, which carry a significant weight in determining the value of alternative distributions, are clearly violated in healthcare. Therefore, addressing the distribution of utilities perceived by individuals in relation to the outcomes of the healthcare they receive (the typical maximand of healthcare cost–benefit analyses) would not be satisfactory. If distributional concerns are to be taken into account, individual and societal utilities with regard to health and healthcare should be reflected more fully in the analysis.

A particularly important form of externality related to health and healthcare is what has been termed ‘interdependent utility’ – that is, the utility accruing to one individual as an effect of someone else’s consumption of healthcare (for a thorough review and discussion of different interpretations of this concept given by philosophers and economists see McGuire and co-workers). Labelle and Hurley have proposed an ‘admittedly crude’ model for measuring this form of externality. They argue that omitting non-user utility from economic evaluation may lead to distortions in resource allocation, whereas by measuring interdependent utilities economic evaluation can address efficiency questions appropriately, as well as some distributional concerns. However, incorporating a measure of interdependent utility into the analysis generates a significant potential for double-counting. In fact, interdependent utility measures may partly duplicate measures of the utility perceived by those who actually receive healthcare and, even more importantly, when a social-welfare function is superimposed the values on which this is based may already have been reflected in interdependent utilities elicited from individuals. Therefore, whereas individual utilities for own consumption of healthcare are not a satisfactory maximand when distributional concerns have to be addressed, the adoption of a broader concept of individual utility that encompasses interdependent utility requires great caution due to the risk of double-counting.

Labelle and Hurley develop their model in relation to an hypothetical preference-based maximand for cost–utility analysis (utils), acknowledging that QALYs are unlikely to reflect utilities. However, in their discussion they mention the possibility of adjusting QALYs to incorporate externalities, such as interdependent utilities in ‘cost–utility’ analysis:

\[ \text{a QALY will be worth more than } 1.0 \text{ for those individuals for whom society more highly values health improvements generating larger than average positive externalities.} \]

Whereas it would appear reasonable to adjust a utility measure (e.g. a measure of willingness to pay) for externalities, the proposed adjustment of a health outcome measure may raise doubts. Externalities are not health outcomes, and whether adopting ‘externality-adjusted quality-adjusted life-expectancy’ as a maximand in economic evaluation may lead systematically to welfare improvements and to an efficient allocation of healthcare resources is questionable. On the other hand, if an extra-welfarist perspective is adopted externalities representing ‘pure welfare’ effects “might be rejected as irrelevant … because no health improvement occurs”, as discussed by Labelle and Hurley. Then, if interdependent utility is deemed irrelevant, why should values regarding alternative distributions matter?

It is important to note that incorporating externalities into economic analysis, particularly interdependent utility, is not in itself sufficient to address distributional concerns. The pursuit of allocative efficiency remains distinct from the pursuit of an optimal distribution, no matter what components of utility are taken into account. Partly in response to a contribution by Hochman and Rodgers, Mishan argues that “efficiency-derived distributions” are impossible, and even if they were possible they would be undesirable. Mishan’s
conclusion is that interdependent utilities should be ignored and distributional judgements should be based on the ethics of the society for which they are intended.

So far the discussion has been focused on the use of measures of health gain (including or excluding externalities) as possible maximands for economic evaluation, because health gain is virtually the only maximand considered in the economic evaluation literature. This limits significantly the set of equity principles that can be used as a basis for addressing distributional concerns through economic evaluation. In particular, only principles regarding the distribution of healthcare according to need (e.g. equal treatment for equal need, equality of marginal met need), when need is defined as capacity to benefit, can be taken into account. Alternative concepts of need (e.g. initial health, expenditure (healthcare) that a person ought to have, expenditure (healthcare) required to exhaust capacity to benefit (as reviewed by Culyer and Wagstaff\(^7\)) may not provide an adequate basis for distributional judgements when the maximand adopted in economic evaluations is health gain, although there have been attempts to adjust measures of health gain to take into account initial health, or disease severity (e.g. Nord and co-workers\(^6\)). A similar inconsistency occurs with regard to alternative equity principles based on access\(^7\) (which may involve the consideration of financial and non-financial costs borne by patients and barriers to use of services), or capabilities,\(^7\) rather than need, when health gain is the maximand.

### Distributional implications of current methods for economic evaluation

In this section we discuss the distributional implications of current methods for economic evaluation with regard, in particular, to cost-effectiveness analysis, which is much more widely than cost–benefit analysis in the healthcare domain. However, the discussion about the principle of health maximisation is largely relevant to both forms of economic evaluation.

### Health maximisation in cost-effectiveness analysis

Here we discuss the distributional implications of the criterion of health maximisation. These criticisms are relevant to cost-effectiveness analysis. We note the potential conflict between the principle of health maximisation and the theories of justice (outlined on pages 10 and 11). We then consider the debate that arose around the application of QALY maximisation to healthcare resource allocation decisions, and raise some of the distributional concerns revealed through surveys of public opinion.

#### Health maximisation and theories of justice

We noted earlier that health maximisation is a particular application of utilitarianism. As such, it potentially conflicts with the principles of ‘maximin’ and egalitarianism (however defined). The group most able to benefit from healthcare per unit cost will not necessarily be the worst-off; inequalities in health between individuals and/or groups may be increased as a result of seeking to maximise health gain. Thus, in general terms, health maximisation may conflict with broad theories of justice.

#### Specific criticisms of health maximisation

The health maximisation decision rule has been severely criticised on grounds of fairness in healthcare decision-making. It has been argued that health maximisation is ageist and that it can also be sexist and racist.\(^{80,81}\) However, this has been disputed by other authors.\(^{82,83}\) Since the application of the principle of health maximisation would systematically favour those with longer life-expectancy (all other things being equal), it is possible that young people, women rather than men, and certain ethnic groups would be favoured. Hence the charges of ageism, sexism and racism. The counter-argument has been that this would not be ageism, sexism or racism, since the grounds for discrimination are not age, sex or race, but health gain per unit cost. But whether or not health maximisation is ageist, sexist and racist, what still remains is the question of whether it is fair to allocate resources to those most able to benefit. Concern has also been raised in this regard for the prospects for disabled people who, all other things being equal, would seem to be less able to benefit from treatment, since the level of health to which they can be returned is likely to be lower according to health status measures. Harris\(^80\) argued that rationing in favour of those most able to benefit imposed the possibility of a “double jeopardy” on certain unfortunate individuals. For example, not only have disabled people suffered the misfortune of being disabled, but then, when requiring healthcare, they find themselves less able to benefit than others on account of the pre-existing disability. Again, however, Harris’s view that this is inequitable has been disputed.\(^{84–86}\)
Kappel and Sandoe, Lockwood and Williams have argued that the application of QALY maximisation may actually discriminate unfairly against the interests of younger people, rather than older people. This would occur when an intervention generates greater benefit per unit cost in older than in younger people. They object to this distributional implication on the basis of the fair-innings argument (see page 12 for an outline of this argument). We discuss in detail arguments concerning the role of age in fair rationing later in this chapter (see pages 23–28).

Culyer and Wagstaff have analysed the distributive implications of various competing definitions of need. Their discussion highlights how health maximisation would not necessarily lead to the treatment of those whose health status, or prognosis without treatment, is most severe. This distributional implication is likely to be regarded by many as an equity problem. A cost-effectiveness league table may give higher ranking to an intervention that provides relatively small benefits to only moderately ill people than to an intervention that provides large benefits to severely ill people. An example of this is given by Hadorn – the State of Oregon’s draft priority list, based on cost-effectiveness analysis, gave a higher ranking to tooth capping than to surgery for ectopic pregnancy.

Distributional implications of health maximisation and public opinion

Surveys of public opinion have revealed that people may in many circumstances disagree with the distributional implications of health maximisation. These surveys are reviewed and discussed in later sections of this report. They have revealed, for example, that people might prefer to choose an intervention which allows the treatment of a greater number of people than the intervention which maximises health gain. This was found in studies by Nord and co-workers and Ubel and co-workers. As already noted, people would seem to have preferences related to age that are not consistent with health maximisation. There is also evidence (see pages 21–23) that people will in some circumstances prefer a more equal distribution of health between individuals, or groups in society, even if this means some sacrifice in the overall health of society.

In summary, there are number of distributional implications of current methods of economic evaluation. To the extent that economic evaluation favours those with greater capacity to benefit, it may systematically discriminate to the advantage of certain groups and against others; it may on occasion discriminate against an individual or group who nevertheless suffers from more severe illness and in favour of an individual or group with less severe illness; it tends to neglect, or may even act to exacerbate, inequalities in health between individuals or groups in society. But whether or not a particular distributional implication arising from the use of economic evaluation is fair or not will depend on the view of equity adopted.

The decision rules of cost-effectiveness analysis

In order to achieve an optimal allocation of resources, cost-effectiveness analysis has to be applied on the basis of the following decision rules:

- a decision-maker is faced with a number of independent healthcare programmes plus clusters of mutually exclusive programmes
- dominated programmes are excluded within clusters (by straight or extended dominance)
- programmes are ranked in terms of their cost-effectiveness ratio (from the lowest to the highest) and selected until the budget is exhausted
- alternatively, a standard price per effectiveness unit can be set and all programmes with a lower cost-effectiveness ratio implemented.

However, it has been pointed out that these decision rules may lead to an optimal allocation of resources only if two key assumptions hold: programmes must be perfectly divisible, and there must be constant returns to scale. These assumptions are unlikely to hold in many healthcare resource allocation problems.

Despite being widely open to criticism, the decision rules of cost-effectiveness analysis do not appear to have any systematic distributional effects, with the exception of discrimination against interventions presenting high cost-effectiveness ratios due to the impossibility of exploiting economies of scale in their production. Examples may include interventions for orphan diseases or interventions aimed at geographical areas with a low population density.

A further undesirable distributional effect of the decision rules of cost-effectiveness analysis, arising in particular when this is based on decision analysis, which typically weights health outcomes by the probability that these may occur and neglects the risk attitudes of the individuals concerned, is that significantly adverse outcomes would appear tiny when they statistically occur with a low
probability. This might, for instance, systematically underestimate the value of potentially life-saving interventions (when the baseline risk of death is relatively small) compared to other interventions, thus contradicting the equity principle described on page 12 as the ‘rule of rescue’.

The equity–efficiency trade-off in health and healthcare

In this section we consider the ‘equity–efficiency’ trade-off as a framework for the incorporation of equity within economic evaluation. Empirical studies of the views of people towards health maximisation and studies of the equity–efficiency trade-off are reviewed.

Conceptual arguments

The equity–efficiency trade-off has been proposed as a framework for the consideration of equity in the allocation of healthcare resources. This approach is based on modern welfare economics, in which there are two dimensions to social welfare: efficiency and equity. In the equity–efficiency literature, equity is mainly defined as ‘equality’.

As an approach to dealing with equity, the equity–efficiency trade-off has two aspects: one normative, the other descriptive. The normative aspect is the proposition that the desirability from a social point of view of distribution of welfare is a function of both the total welfare and the distribution of this welfare between people. The descriptive aspect is the proposition that making a trade-off between equity and efficiency when asked to make distributive judgements, is something that people just do. They would not, for example, apply a principle such as maximin or utilitarianism. They rather act to strike a balance, according to their preferences, between two desirable, but conflicting, objectives. How this balance is achieved is determined by each individual. The resulting social ranking of different programmes will, therefore, depend on the preferences of the person, or group of people, sitting in judgement. The preference for equity is characterised as ‘inequality aversion’. Thus, despite having an aversion to inequality, a society may accept an increase in inequality, for example, in life-expectancy between social classes, in return for a compensating increase in the overall life-expectancy of the community. Conversely, a society may accept a decrease in overall life-expectancy in return for greater equality between social classes.

The equity–efficiency trade-off in the health economics literature

Wagstaff argued that the equity–efficiency trade-off has both theoretical appeal and practical applicability. He discusses various concepts of equity (equal treatment for equal need, equality of access, equality of health, equity as choice) and assesses each against a three-point criterion:

- Can the definition of equity provide a basis for determining an equitable allocation of resources?
- If so, is there a conflict between equity and efficiency?
- If there is a conflict, can the definition of equity be captured simply by weighting QALYs appropriately in the objective function?

He argues that equality in health is the only principle on offer that provides a clear definition of equity. However, the pursuit of equality at any cost in terms of efficiency would be undesirable. He suggests retaining the concern for equality, but balancing this against efficiency. Thus, this approach, if it captures our views on equity, could be a way of retaining but modifying QALY maximisation to incorporate equity.

Dolan develops more formally the approach described by Wagstaff. He develops the theory of the health-related social-welfare function (HRSWF). These social-welfare functions are defined in relation to what Dolan calls ‘health-related social welfare’. We find in a footnote that it is the distribution of health-related utility rather than health itself that Dolan sees as the ultimate outcome at stake in a HRSWF. He explains:

> Although many authors define the social-welfare function over the utility space generally, because the concern here is with health-related utility, for consistency the term ‘health-related’ social-welfare function will be used throughout this paper [our italics].

Dolan offers QALYs as a measure of this health-related utility. There are a variety of possible HRSWFs. One of these incorporates a trade-off between equity and efficiency in which the choice between alternative distributions of health across society is made in order to strike a balance between these two objectives. Alternative HRSWFs include, for example, the Rawlsian HRSWF and the utilitarian HRSWF. Under the Rawlsian HRSWF, the choice between alternative distributions of health across society is made in order to maximise the health of the least healthy individual or group. Under a
utilitarian HRSWF, the choice between alternative distributions of health across society is made in order to maximise the health of the entire society. It is up to empirical work to reveal which of the possible HRSWFs represents the views of the population.

Williams proposes the ‘fair-innings’ argument to support incorporating a concern for equality into cost-effectiveness analysis. He suggests eliciting weights from members of the public to reflect the extent to which people are prepared to sacrifice efficiency for improvements in equity. He demonstrates how this might be done using hypothetical values.

In summary, health economists have argued that a concern for equality can be incorporated in cost-effectiveness analysis through the application of equity weights. The survey evidence regarding whether people are willing to trade-off efficiency for equality is reviewed below.

Weights for severity
The common belief that healthcare should be distributed according to need is interpreted by many to mean according to severity of pretreatment condition. Severity of illness will receive no weight under a health maximisation decision rule: someone with moderate illness, but a greater capacity to benefit per unit cost, would be preferred to someone with more severe illness, but a lower capacity to benefit per unit cost. But ‘severity’ needs to be defined in order to form the basis for a valuable equity principle. Should it be measured in terms of present health status or include past and/or future expected health? Nord and co-workers define severity as expected utility in the absence of treatment, where expected utility is measured by expected QALYs. This view takes prognosis into account, as well as immediate health status, when assessing severity. Nord and co-workers suggest taking account of severity via weighting the cost-effectiveness ratio.

Empirical evidence
Here we present the methods and results of two types of empirical study: studies which have investigated whether people use the principle of health maximisation when asked to make hypothetical rationing choices; and studies which have quantified the weight placed on equity through experiments aimed at demonstrating evidence for or against a trade-off between equity and efficiency.

Are people ‘health maximisers’?
A survey carried out in Norway by Nord asked respondents to choose between two people, each of whom was suffering from a life-threatening condition, where one could be returned to full health and the other to less than full health. The sample consisted of workers at the National Institute of Public Health in Oslo. Of the respondents, 79% were not prepared to discriminate, preferring that the two patients be treated in the order in which they were admitted to hospital; 15% only gave priority to the patient with the better expected outcome. A second questionnaire asked people to make a person trade-off. They were asked to choose between two units, A and B. Unit A would save 10 lives and return these people to full health. Unit B would save lives, but these would be at less than full health: “a life with moderate pain and dependency on crutches for walking”. They were asked how many people would need to be saved in unit B to be equivalent to 10 lives saved in unit A. A second version of this question was given in which the benefits of treatment of patients were purely in terms of quality of life. The results of these person trade-off exercises were consistent with the findings of the first questionnaire, revealing a preference that expected health outcome should not form a basis for discrimination.

In an Australian survey of public views towards a range of rationing decisions, Nord and co-workers asked respondents whether they agreed that among patients suffering equally, priority should be given to those who would be helped most from treatment. There was a slight majority in favour (52.8% agreed; 47.2% disagreed).

In another Australian survey, Nord and co-workers asked people to compare treatments of different duration using the person trade-off method. They were asked questions such as: “Consider two projects, A and B. In project A, the lives of 10 people could be extended for 10 years in normal health. In project B, patients’ lives would be extended for 20 years. How many patients treated in project B would be equivalent to the 10 people treated in A?” The sample of volunteer interviewees was drawn from respondents to a larger survey of attitudes to healthcare rationing amongst people in Melbourne, Australia. The results indicated that people were willing to sacrifice a certain amount of health gain in order to treat more people.

In a US study, Ubel and co-workers studied attitudes towards equity and efficiency in choosing
between two screening tests for cancer. One of these tests was more effective and saved more lives. But it was also more expensive. Given budget constraints, the more effective test could not be offered to everyone in the population who might benefit. The selection of people to be offered the more effective test would be made randomly. The outcome was that 56% of prospective jurors (members of the public), 53% of medical ethicists and 41% of experts in medical decision-making chose to offer the less effective test to the whole population. In doing so, they were prepared to sacrifice the 100 extra lives that would be saved by offering the more effective test to half the population.

Abellan-Perpinan and Pinto-Prades\textsuperscript{94} conducted a study in Spain to test the extent to which respondents used health state after treatment as a reason for discrimination. They conducted two experiments and found in both that there was considerable support for not discriminating on the basis of health potential. However, some support was also found for discriminating on this basis. The authors concluded that “… models that combine efficiency and equity should clearly play a more important role in priority setting”.

**Weights to reflect the equity–efficiency trade-off**

We describe here three empirical studies which have sought to derive a measure of the trade-off between equity and efficiency. Two of these studies are from Sweden\textsuperscript{95,96} and one is from the UK.\textsuperscript{91} A fourth study\textsuperscript{97} has investigated the equity–efficiency trade-off in the context of choosing programmes with different effects on equity between occupational groups. This work is described and discussed later in this chapter (see pages 21–23).

Johannesson and Gerdtham\textsuperscript{95} asked 80 economics students to choose between two societies, A and B, which differed in terms of the distributions of future health between two social groups. Society A contained a greater average remaining lifetime health, and society B contained a narrower distribution of remaining health. The students were told to imagine they would belong to one of these societies, but would only find out which once they had chosen between them. They had a 50% chance of belonging to either. The sample was randomly divided to receive two versions of society A and four versions of society B. The two versions of society A allowed the relative inequality between social groups 1 and 2 to vary. The four versions of society B allowed the trade-off between equity and efficiency to vary (this trade-off being the ratio of the loss in QALYs in group 1 to the gain in QALYs in group 2). The trade-off was allowed to vary between 0 and 1. A statistical model was developed to test two hypotheses:

- an increased marginal trade-off would decrease the probability of choosing society A
- an increased inequality in society A will decrease the probability of choosing A.

The outcome variable was the probability of choosing society A. They also calculated the median trade-off. The results were consistent with the first hypothesis: as the gain in QALYs for group 2 in society B increased relative to the loss in QALYs in group 1, the probability of choosing B increased. The second hypothesis was not supported by the data. The marginal trade-off was 0.45, meaning that those in the survey were willing to sacrifice 1 QALY from group 1 (those better off in terms of remaining lifetime health) to gain 0.45 QALYs in group 2.

Andersson and Lyttkens\textsuperscript{96} also used the ‘veil of ignorance’ approach to design scenarios for their survey of preferences for equity. They investigated two issues: they looked for evidence to support the equity–efficiency trade-off; and investigated whether the theoretical difference between risk (known probabilities) and uncertainty (unknown probabilities) affected choice from behind a veil of ignorance. The sample comprised 225 economics students. The students were asked to choose between two societies that each consisted of two groups of people. The distributions of lifetime health (measured in life-years) between group 1 (the fortunate) and group 2 (less fortunate) differed between the two societies. In both societies group 1 enjoyed greater life-expectancy than group 2, but in society B the fortunate lived a shorter life than in society A and the unfortunate a longer life. Half of the students were asked to choose from a position of known probabilities, and half were to choose from unknown probabilities. As in the study by Johannesson and Gerdtham,\textsuperscript{95} the trade-off between groups 1 and 2, and the inequality in society A, were varied.

The results of the study showed no difference between those who were given probabilities and those that were left in uncertainty (the proportion choosing society B was 73% among those not given probabilities and 69% among those who were). However, both the size of the trade-off and the inequality in society A affected the responses. In a statistical model with the probability of choosing society B as the outcome measure, the trade-off was
found to be highly significant and the inequality in society A to be reasonably significant \((p = 0.074)\). The median marginal trade-off (the point at which half of the students preferred society A and half preferred society B) was 0.35 for the more equal distribution and 0.11 for the more unequal distribution.

Dolan\(^91\) has reported the results of a small pilot study to investigate how people value health gain in individuals with differing initial health states. The sample consisted of 35 students, who were asked to choose between treating individual \(i\) with initial health status of 0.2 and an individual \(j\) with initial health status of 0.4. Health gain from treatment was fixed at 0.2 (health status after treatment 0.4) for individual \(i\). Respondents were asked at what level of health gain for individual \(j\) they would be indifferent between treating \(i\) and \(j\). The responses showed that moving one person from 0.2 to 0.4 was equivalent to moving another person from 0.4 to 0.8. This experiment also revealed that some in the sample did not judge any level of increased health gain to the more fortunate person to be equal in value to a gain of 0.2 to the less fortunate person.

**Discussion**

When interpreting and comparing survey results, it is important to be aware of the influence on responses of how questions are framed. However, given the variety in the framing of the empirical studies described above, some general conclusions can nevertheless be drawn. It seems clear that the principle of health maximisation is unlikely to provide a universal principle commanding public support in all circumstances. The findings of Nord and co-workers\(^92,93\) and Abellan-Perpinan and Pinto-Prades\(^94\) suggest significant opposition to the simple application of health maximisation. More convincing evidence of this was found from the studies involving the possibility of trade-offs. The person trade-offs reported by Nord and co-workers\(^97\) and Ubel and co-workers\(^98\) revealed that many respondents were willing to make sacrifices in health gain in order to increase the number of people that benefit from an intervention. The studies by Johannesson and Gerdtham\(^95\), Andersson and Lyttkens\(^96\) and Dolan\(^91\) were designed to quantify the trade-off, should people make one, between total health in a population, or groups of individuals, and the distribution of health between individuals. Each of these studies found evidence of such a trade-off.

Both Swedish studies found choices to be sensitive to the trade-off between health gain and the degree of inequality between the two social groups. Johannesson and Gerdtham\(^95\) estimated a willingness to sacrifice 1 QALY from group 1 (those better off in terms of remaining lifetime health) to gain 0.45 QALYS in group 2. This is equivalent to a relative weight for health gains in group 2 compared to group 1 of 2.2 (1/0.45). Andersson and Lyttkens\(^96\) found choices to be sensitive to the degree of inequality in society A. When the distribution was more equal, the trade-off implied a weight of 2.9 (1/0.35); when inequality was greater, the implied weight was 9 (1/0.11).

Dolan\(^91\) tested whether people would trade-off health against equity when choosing between two individuals who differ with respect to initial, pretreatment health status. He found that the health gain available to the person in the more severe health state was valued twice as highly as health gain in the person with less severe pretreatment health status.

In conclusion, these quantitative studies provide evidence that people will, in certain contexts, trade-off equity and efficiency, and that such experiments might be used to derive equity weights. We have evidence of trade-offs in four contexts:

- when choosing between healthcare programmes which differ with respect to overall health gain and the numbers of people who benefit\(^97,98\)
- when choosing between two groups which differ with respect to remaining lifetime health\(^95\)
- when choosing between groups which differ with respect to total lifetime health\(^96\)
- when choosing between two individuals with differing pretreatment health status (severity).\(^91\)

In terms of the HRSWFs discussed by Dolan\(^91\) (see page 16) this evidence would seem to reject the utilitarian and Rawlsian HRSWFs. Later in this chapter (see pages 30–34) we consider further the evidence from these quantitative studies when discussing whether equity weights provide a way forward for the incorporation of equity within economic evaluation.

**Equity weights and economic evaluation**

**The concept of equity weighting**

Equity weights provide a method for formally incorporating concepts of fairness into economic
analysis. The equity weights are numbers that express the relative importance of the relevant equity concepts. They thus represent the quantification of abstract concepts in order to operationalise and utilise these concepts in formal analysis.

An equity weight in cost-effectiveness analysis expresses the extent to which society is prepared to sacrifice health gain in the pursuit of fairness. The greater the equity weight, the more health gain a society is willing to sacrifice to achieve improved fairness. Therefore, the social preference between various healthcare programmes will depend not only on the magnitude of the equity weights, but also on the relative cost-effectiveness of healthcare programmes. The more efficient the less equitable programmes are, the larger the equity weights will have to be to secure a higher ranking for the more equitable programmes.

An underlying assumption necessary to the development and application of such weights is that the concepts of equity and the concept of efficiency can be traded off against each other. The application of weights is thus used to effect a balancing of conflicting, but commensurable, objectives when making morally complex resource allocation decisions. However, there may be constraints and limits to this weighing-up process. For example, such a constraint may be that all patient groups should be offered at least some level of care (subject to that care being effective), even if greater health gain overall can be derived from denying certain groups care. Thus the equitable distribution of healthcare resources might be achieved through a mixture of equity-based constraints and equity-weighted claims where competing principles are to be traded off.

Sources of equity weights
Equity weights may be derived from two main sources:

• the views of a (representative) sample of the population
• the views of decision-makers.

The views of the population
The health economics literature has tended to favour the views of a (representative) sample of the population over the views of decision-makers. In an investigation of the views of a population, experiments are done using samples from the population (ideally representative samples), in which the respondents are asked to make choices designed to reveal the extent to which they are willing to make trade-offs between equity and efficiency, or between numbers of people treated. There are a small number of empirical studies of this type. This initial work has indicated that people are willing to make trade-offs between alternative principles.

A less rigorous application of this approach is to base the weights on those implied by public opinion surveys, rather than deriving them directly through choice experiments. Where this method is adopted, the actual numbers are assigned by the analyst rather than derived directly from the public. The age-weighting function used to weight disability-adjusted life-years (DALYs) for the World Development Report 1993 was in part justified by appeal to surveys of public opinion. DALY age weights were based on the view that the health of people at more productive times of life should be weighted more highly than the health of people at less productive times of life. This was supported by reference to studies of public opinion, such as those by Busschbach and co-workers and Lewis and Charney.

Expert or decision-maker opinion
There are two approaches to obtaining equity weights from decision- or policy-makers:

• to elicit their preferences directly through some kind of survey
• to infer their preferences through analysis of past policy decisions.

We are aware of one empirical study to have surveyed policy-makers for the purpose of deriving weights. Lindholm and co-workers asked politicians involved in the process of allocating healthcare resources in Sweden to make choices between hypothetical healthcare programmes, in order to reveal whether, and to what extent, they were prepared to trade-off efficiency for equity.

However, we are not aware of examples in the health economics literature of attempts to infer equity weights implicit within actual policy decisions. Government policy statements may provide a clue to determining government objectives. But it might be possible to infer the preferences of government through analysis of the distribution of benefits and costs resulting from government policy. It has even been suggested that if enough choices are observed it might be possible to reveal the precise preference patterns of government. Basu has described various models of government decision-making under which distributional weights could be inferred. These models differ in the extent to which government is viewed as possessing ‘rationality’. The fully rational government acts in accordance with well-defined
objectives, unhindered by political compromise, with full knowledge of the consequences of alternative policy options, and has the ability to implement policy in accordance with its objectives. However, objectives may be unclear and power may be diffused such that there are unintended effects. In addition, the evaluation of options is costly. Thus, attempting to infer the preferences of government and other health policy decision-makers (Health Authorities) is a complex undertaking likely to raise questions as much as provide answers. The absence of analyses of governmental health policy decisions may be due to difficulties related to identifying the effects of policy, dealing with multiple decision-makers and knowing whether the effects of policy decisions are in fact intended.

How equity weights are applied

Under traditional cost-effectiveness analysis, health gains are weighted equally, irrespective of to whom they accrue. For example, one QALY is valued as one QALY, irrespective of characteristics such as health status, age, ethnicity or social class, that might be relevant to equity. The application of equity weights implies weighting health gains differently, depending on the (equity relevant) characteristics of the recipients. Equity weights would be applied as an adjustment factor to a cardinal measure of health gain to take account of such characteristics. The resource allocation task is then to maximise the equity-weighted sum of health gains. Such a process of adjustment may influence resource allocation based on economic evaluation by changing the ranking of cost-effectiveness ratios in a league table (indeed, equity weighting would be redundant if the ranking of cost-effectiveness ratios remained unchanged after the application of equity weights).

Equity across social groups with different levels of health

Conceptual arguments

A limited health economics literature has developed around the issue of equity weighting on the basis of social categories such as social class.69,97,102–104 Two important questions in relation to this issue are:

- Inequity in terms of what?
- Inequity between whom?

In answer to the first of these questions, Lindholm and co-workers97,101,104 and Williams69 consider that it is inequality in health that should be the focus of equity weighting in healthcare. This perspective on equity considers the distribution of health across (subgroups of) the population. The objective of equity weighting is to correct for unacceptable inequalities in health that might exist between various social groups. The development and application of weights on such a basis would imply that an individual’s claim on healthcare resources might be increased or decreased depending on the relative health experience of the social group(s) to which that person belongs.

However, Le Grand102 has discussed the possibility that it might be inequality in non-health factors, such as economic or political position in society, which should be the basis for equity weighting in healthcare, rather than inequality in health. He asks:

... might equity require that health care for the poor, for ethnic minorities or for women, for instance, be given priority over health care for the better off, for the white population or for males – on the grounds that this offered them some compensation for their disadvantage?102

In principle a weighting scheme might be developed that gives greater weight to units of health gain in disadvantaged individuals than to units of health gain possible in people from more privileged positions in society.

In terms of the second question (Inequity between whom?), relevant social groups might include, for example, socio-economic classes, ethnic groups, men and women, and behavioural risk groups (e.g. smokers and non-smokers). Williams69 has suggested that social class might be one basis for the development of weights to reflect equity considerations. Recent research in the UK has shown a social-class gradient in health, where those in higher social classes enjoy greater life-expectancy than those from lower classes.105 Williams suggests that this is an unfairness which might be a basis for the derivation of equity weights. Inequality in health between occupational groups has also formed the basis for an empirical study to derive equity weights (Lindholm and co-workers97 derived a quantitative measure of the weight placed on equity between blue- and white-collar workers by politicians in Sweden).

Empirical evidence

In this section we review results from surveys of public views of how characteristics such as sex, ethnicity and social class should be taken into account when setting healthcare priorities. Only
one study provides a quantitative assessment of the weight given to equity.

Charney and co-workers asked respondents to make choices for life-saving treatment between people on the basis of occupational status, marital status, sex and lifestyle. They found that approximately 50% of their study subjects were prepared to prioritise healthcare on the basis of occupational group. When choosing between an unemployed and an employed person for life-saving treatment: 33.1% favoured priority to the employed person and 15.6% the unemployed person (the rest expressed no preference); 21.5% favoured a director over an unskilled worker versus 28.5% the unskilled worker over a director; and 34.5% favoured saving a teacher to a lorry driver, while 15.2% preferred the lorry driver. There were clear majorities in favour of prioritising married people over non-married and those with healthy lifestyles over those with less healthy lifestyles (smokers versus non-smokers, high alcohol versus low alcohol, poor diet versus inherited risk). A majority gave equal priority to men and women. Amongst those that did choose, there was a preference for women.

Bowling reported that 42% of respondents (from a representative sample of the UK population) either agreed or strongly agreed that lower priority should be given to those that contribute to their illness through smoking, obesity or excessive drinking.

Dolan and co-workers report the results of a focus-group study that took place in North Yorkshire involving 60 subjects. Views were tested prior to and after discussion. Respondents were asked to indicate whether a particular group should have ‘much more’, ‘more’, ‘the same’, ‘less’ or ‘much less’ priority than others. The following results represent views after discussion. Concerning socioeconomic status, those who expressed priority did so for poorer people (e.g. 23% gave lower priority to the rich, 0% higher priority; 8% favoured higher priority for those with low education versus 2% who gave lower priority). Concerning lifestyle, those who expressed priority favoured lower priority for smokers, heavy drinkers, people with unhealthy diets, those who rarely exercise and illegal drug users (the latter were given lower priority status by 43% of study participants). However, in each of these cases, more than 50% expressed no priority. There was negligible desire to prioritise on the basis of sex, whether a person was white or not, or marital status. Finally, 35% favoured higher priority to the disabled.

Kneeshaw reviewed a number of UK surveys of attitudes to rationing. He cites a Gallup survey conducted in June 1994 in which people were asked who should go to the back of the queue for hospital treatment (cited in New, Table 5.9). The majority of respondents were against de-prioritising on the basis of smoking, heavy drinking, bodyweight or overwork. A second Gallup poll in August 1994 (cited in New, Table 5.9) asked a similar question and found a closer outcome: 41% agreed, 49% disagreed.

An Office of Population Census and Statistics survey (also cited in New, Table 5.9) found a similar result: 42% agreed that those who contribute to their own illness should have lower priority; 43% disagreed.

Lindholm and co-workers have reported the methods and results of a pilot study in which politicians responsible for healthcare in the Swedish county of Alvsborg were sent questionnaires asking them to choose between hypothetical programmes aimed at reducing risks of death from myocardial infarction. Two social groups were defined: blue- and white-collar workers. Health was measured in terms of prevented deaths. The first programme saved more lives overall (100) but made no change to equity (blue-collar workers remained with a death rate from myocardial infarction 50% higher than that of white-collar workers); the second programme improved equity between blue- and white-collar workers by equalising their death rates from myocardial infarction, but saved only 90 lives overall. The survey was sent to 68 people. In total 76% responded. Overall, 15 gave priority to the more equitable programme, 27 gave priority to the more efficient programme and six were indifferent. However, specific equity weights were not derived.

Lindholm and co-workers have reported a study to derive a quantitative assessment of equity. A questionnaire was mailed to politicians responsible for healthcare in 10 counties; 449 questionnaires were returned. The politicians were given the same scenario used in Lindholm and co-workers and were asked to choose between a health maximisation programme and one of three randomly assigned equity programmes. These equity programmes varied in the sacrifice of lives saved in the white-collar workers necessary to reach equality in death rates between the white- and blue-collar workers.

The results showed that as the sacrifice of efficiency increased so did the proportion of survey participants preferring the more efficient programme. It was estimated (through linear
interpolation) that the median willingness to sacrifice efficiency for equity was 15% of health gains. They defined a measure $E$ of inequality aversion as the ratio of an equal distribution of health to an unequal distribution of health with a higher mean, where society is indifferent between the two distributions. They presented a formula for equity-adjusted years of life saved:

$$EYLS = YLS[E + (1 - E)g]$$

where $EYLS$ is the equity-adjusted years of life saved, $YLS$ is the years of life saved, $E$ is a measure of inequity aversion ($E = 1$, no inequity aversion; $E < 1$, inequity aversion), and $g$ is a measure of the extent to which equality is achieved (i.e. the percentage reduction in relative risk between the two groups).

**Suggested equity weights**

We discuss here what kinds of weights, if any, are supported by the empirical studies described above. The discussion is in two parts: first, we consider weights relating to socio-economic status; and, secondly, we look at weights relating to lifestyle choices.

**Socio-economic status**

Charny and co-workers found that roughly half their sample were prepared to make rationing choices for life-saving treatment based on occupational status. However, it seems unlikely that this was based on a concern for the distribution of health. The study design did not make respondents aware of health differences between occupational groups. It is therefore likely that judgements were made on the basis of perceived social value of alternative occupations, rather than notions of equity in relation to health inequality. However, it indicates that considerations of productivity might be important to many people when making rationing judgements. In contrast, Dolan and co-workers found less of a preference for prioritising on the basis of socio-economic status. This difference is likely to be accounted for by differences in study methodology, in the characteristics of respondents and the way in which questions were framed. In particular, the questions posed by Charney and co-workers were more demanding in that they required a life or death decision be made between people from specific occupations, rather than a broad statement of preference for one group compared to another.

Do these studies suggest a particular pattern of weights? It is not clear from these studies what motivated those who were willing to prioritise on the basis of socio-economic status. Where a preference was expressed for the person with lower socio-economic status, was this based on attitudes towards health inequalities or general socio-economic inequality? To what extent are answers based on judgements about productivity and value to society? How did respondents perceive the socio-economic status of different occupations, or their relative health status? These two studies were not designed to be able to provide such evidence. They are preliminary investigations of public preferences.

However, the two Swedish studies investigated more closely the issue of equity between socio-economic groups with differences in health. Extending their earlier study, Lindholm and co-workers offered a choice of equity programmes to their respondents, each requiring a different level of sacrifice of overall health gain. They found that as the sacrifice of efficiency increased the proportion preferring the more equitable programme decreased. This study supports equity weights in favour of blue-collar workers with lower health status than those in higher socio-economic groups. However, it is a Swedish study, and therefore is not necessarily applicable to the UK.

**Lifestyle**

There is currently no empirical study which has been designed to quantify the extent to which the general public may wish to prioritise (or de-prioritise) those who follow particular lifestyles. The attitude surveys report a mixed picture. However, overall, the balance of these surveys suggests that most people would not support prioritisation on the basis of risk behaviour such as smoking or heavy drinking.

In conclusion, these studies have focused on whether there is a basis in public opinion for rationing healthcare based on factors such as socio-economic status and lifestyle choices. It is not clear from these studies what principles people might be applying when making their choices. Variation in design makes comparison between studies difficult. Only one study has quantified preferences in this area, and this work was not based on a UK sample. There is currently no evidence to support equity weighting of this type in the UK.

**Equity across age groups**

The issue of how a person’s age ought to be treated in making choices between people in distributing scarce healthcare resources forms a prominent
debate within the literature. Radically opposing views have been expressed on the appropriate role of age in making a fair allocation of scarce healthcare resources. In this section we review the various bases on which age-related priority setting might be justified, and discuss the arguments advanced in defence of each. We then review surveys of public opinion in relation to the role of age in setting priorities for healthcare, and ask what basis exists for the application of age-related weights.

**Conceptual arguments**

Three arguments might be made for valuing unequally health gains to people of different ages when making resource allocation decisions. Below we discuss these arguments and the results of some surveys of public opinion of how age should influence the value of health gains. In particular we review studies that have quantified the extent to which the public might judge health at some ages to be more or less valuable than health at other ages.

**People value their own health differently at different ages**

According to this view, the value of health varies with age because being healthy happens to be more valuable to a person at certain times of life than at others. It has been suggested that when first starting education as a child it is especially important to be healthy. This might be because this is a crucial time that affects future development and opportunities for a person. Others have suggested that becoming a parent of young children is a similarly important time of life. Less valuable times might include stages nearer the end of life.

**The debate**

Some have argued that the value of being healthy to an individual varies with the age of that individual. For instance, Kappel and Sandoe argue that the value of health at a particular time of life depends on the ‘life-plans’ of individuals. On this view, the utility of health depends on the extent to which people need health to achieve their aims and goals. If older people have achieved most of their aims and goals, the utility to them of health is diminished. In contrast, if a younger person is in the process of fulfilling important, high utility activities (e.g. establishing a career), health will be valued more highly. In general terms, it seems plausible that to be unable to walk at, for example, the age of 25 years is worse for someone’s utility than being unable to walk when 75 years old. If QALYs are intended to capture utility, then they might need to be adjusted to incorporate a preference for health that is related to time of life.

However, Kappel and Sandoe add the proviso that it is not always the case that the value of health will necessarily be higher to the younger individual. They write: “And if old people start to expect more from life the health-states needed to fulfill these expectations will also become more useful”.

Harris, on the other hand, has argued against using personal preferences as a basis for distributing resources between people. He says that the fact that people may prefer something (e.g. good health status at one time in life rather than at another) does not imply that this preference should be used to make allocative decisions between people. The fact that an individual might prefer to be healthy when aged 25 than when aged 75 does not imply that he or she is committed to the view that the health of the 25 year old is of greater social value than the health of the 75 year old. Harris is here implicitly drawing upon the distinction between efficiency, as the maximisation of utility, and equity.

**A person’s health has value to others in society that varies with a person’s age**

The value of a person’s health to other people in society might be considered to be greater at certain times of life than at others. For example, the health of parents is important to the development of children; the health of those who are economically active is valuable to the whole of society. By implication, the health of those without dependants or those who are not economically active may be less valuable to others in society. Thus, there is a case in terms of wider social benefits for the prioritisation of the health of people at ages when they are likely to be of relatively high value to society (e.g. when in the role of parent or in employment).

**The debate**

It has been argued that each person’s productivity and value to others in society (either to their immediate family or to wider society) might be a function of age, where this value begins at zero, rises to its highest point during adulthood and falls in old age. This pattern is intended to reflect how each person’s contribution to the welfare of others changes through life, beginning with dependency on others in early life through to independence and support of others during adulthood, and returning to dependency again in later life.

The calculation of DALYs used by the World Bank incorporated an adjustment for this aspect.
of the value of health. Murray and Acharya write in defence of this age weighting of the DALY: “Even if every year of life has the same intrinsic value to the individual, we may be ‘induced’ to attach greater importance to years of productive adult life.” They go on to argue that, in general, we require support as children and may require care again near the end of life. As adults, however, we are likely to be needed to support others, both economically and emotionally. They conclude: “The well being of some age groups, we argue, is instrumental in making society flourish; therefore collectively we may be more concerned with improving health status for individuals in these groups.”

Murray and Acharya are here proposing that non-health benefits to others should be used to help set healthcare priorities. But should impacts on the lives of others (e.g. on dependants) determine treatment priorities between people in need of healthcare? Such impacts have been called ‘side-effects’. One concern of including side-effects is how to confine them solely to divisions of the population according to age. The extra benefits of one person’s health over another’s might be invoked as an argument to support priority setting on the basis of other characteristics, such as income, social class or education. Murray and Acharya are aware of this implication. They respond by arguing that age does not discriminate between the lives of different people, but rather between periods of the life cycle of each individual. In contrast, characteristics such as income discriminate between individuals. A counter-argument to this, that we are not in fact the same person at different stages of our lives, is discussed by Parfit.

**Equity in health means equality in lifetime health**

Those who are older have enjoyed more healthy lifetime than those who are younger. Therefore, it could be considered unjust to value gains in health to older people equally with gains in health to younger people, since the inequality in lifetime health enjoyed by older and younger people would only increase. Since we should aim at equality in lifetime health, health gains to the young should be valued differently from health gains to the old.

**The debate**

There has been, and there still is, a lively debate about the implications of age as an equity dimension in the allocation of healthcare resources. The principle of QALY maximisation has been criticised for discriminating unfairly against both the old and the young. In this section we describe these arguments.

Harris argues that QALY maximisation is ‘ageist’ on the basis that it implies unfair discrimination against older people. He makes an argument from the principle of equality. This requires that people are treated with equal concern and respect, which implies treating all people’s fundamental interests equally. Each individual who faces death if denied treatment, and who wishes to go on living, has an equal moral claim because each faces losing the rest of their life. Harris argued that the QALY calculation will tend to favour younger people since they have a greater life-expectancy, and consequently will violate the principle of equality.

Others have argued that, contrary to being ageist in the sense of discriminating unfairly against the interests of older people, QALY maximisation may in fact discriminate unfairly against the young. This argument is also made from a principle of equality. This principle is the view that equity in health implies equality in age at death (or, perhaps more precisely, equality in lifetime health) (see, e.g., Lockwood, Kappel and Sandoe and Williams). The argument is as follows: QALY maximisation has the potential to favour the young or the old (it depends on relative capacity to benefit per unit cost). On the assumption that older people have enjoyed more healthy years of life than those younger than them, then favouring the old, where they have higher health gains per unit cost, would only tend towards increasing this inequality. Therefore, QALY maximisation has the potential to discriminate unfairly against the interests of the young. This view, that the young should be prioritised in order to, as far as possible, equalise total lifetime health, has been called egalitarian ageism.

**Empirical evidence**

Here we consider evidence in relation to what people in general think about the role of age in determining priorities for healthcare. We briefly summarise the key results of the surveys we were able to retrieve. The implications of these studies for equity weighting will be discussed in the following section.

Williams was interested in the value to individuals of health at different ages. He analysed responses from 377 people to a section of the York Health Evaluation Survey conducted in 1985, and found that the life stages ‘when bringing up children’ and ‘as infants’ were classified as the most important by respondents asked to rank ten life stages in order of importance (these categories received 32.9% and 27.3% of the first-place ranking).
Bowling\textsuperscript{107} conducted an interview survey of a random sample of the UK population. Participants were asked to rank 12 healthcare programmes in order of priority. People tended to give highest priority to ‘treatments for children with life-threatening illnesses’ and lowest rank to ‘treatment for people aged 75 and over with life-threatening illness’.

In contrast, Nord and co-workers\textsuperscript{93} in an Australian survey, found quite a strong level of support for non-discrimination on the basis of age. When choosing who should receive life-saving treatment, the young or the old, 41.9% favoured equal priority, 40.5% less priority to the very old and 17.6% favoured priority to the young. For treatments to improve quality of life, 75.6% gave equal priority to young and old, 21.3% favoured the young, and 2.9% the old. Finally, when choosing between a young child and a newborn, 54.7% gave equal priority, 44.2% favoured the young child and 1.2% the newborn.

Other studies have required respondents to make direct rationing decisions between people of different ages. Lewis and Charny\textsuperscript{100} asked respondents (721 people selected randomly from the electoral register for Cardiff) to make three choices over to whom to give life-saving treatment. They found a preference for a 5 year old over a 70 year old (a ratio of 84:1 of respondents who expressed a preference), a 35 year old over a 60 year old (a ratio of 14:1) and an 8 year old over a 2 year old (a ratio of 5:3).

In a Dutch study, Busschbach and co-workers\textsuperscript{99} interviewed study subjects about whom to prioritise for treatment to improve quality of life. Choices were required between people at the ages of 5, 10, 35, 60 and 70 years. Total lifetime health for each hypothetical person would be the same, but the age at which they would experience ill-health was not. Thus the experiment aimed to elicit the value of health at different ages. In general, the results showed health to be preferred when younger. The exception to this was the finding that health at age 10 was valued more highly than health at age 5.

The weights were 1.2, 1.5, 1.0, 0.7 and 0.7 for ages 5, 10, 35, 60 and 70 years, respectively (as cited in Tsuchiya\textsuperscript{115}).

Cropper and co-workers\textsuperscript{116} carried out a telephone survey of nearly 800 people in the USA to investigate preferences for saving the lives of people of different ages. They used the person trade-off method to elicit values for saving lives of people aged 20, 30, 40 and 60 years. The number of life-years gained was allowed to vary. They found that people placed more weight on young people than that accounted for by longer life-expectancy. Saving one 30 year old, one 20 year old and one 40 year old was equivalent to saving eleven, eight and seven 60 year olds, respectively. The implied weights are 0.7, 1.0, 0.6 and 0.1 for ages 20, 30, 40 and 60 years, respectively.

Nord and co-workers\textsuperscript{87} conducted interviews in which respondents were asked to make choices between treatment programmes (both life extending and health improving) for people in four different age groups (10, 20, 60 and 80 years old) using the person trade-off method. The study subjects were 176 volunteers who had already taken part in a larger postal survey. Health gains were equalised between the age groups. The study took place in Australia. The results showed that the younger groups were preferred to the older patient groups. The weightings for ages 10, 20, 60 and 80 years were 1.1, 1.0, 0.4 and 0.1, respectively. These weights were approximately the same for both life-extending and health-improving treatments.

Johannesson and Johansson,\textsuperscript{117} in a Swedish study, asked people to choose between an older person (50 or 70 years old) and a younger person (30 years old) when both required life-saving treatment, but only one could be treated. Equity weights were derived for lives saved, life-years gained and QALYs gained. The weights were calculated undiscounted and discounted at 5%. The weights (undiscounted) for lives saved at the ages of 30, 50 and 70 years were 1.0, 0.13 and 0.02, respectively. For life-years gained, the weights were approximately the same (1.0, 0.22 and 0.1, respectively). The weights for QALYs gained were virtually identical to those for life-years gained.

The review by Tsuchiya\textsuperscript{115} has reported two Japanese studies to have investigated age-related preferences. Tamura and co-workers\textsuperscript{118} found the majority of respondents to a postal survey gave priority to the younger person when asked who should receive emergency treatment first. They were asked to make the following comparisons: 20 versus 60; 20 versus 80; 5 versus 0; 60 versus 80 and 5 versus 20 years. Tsuchiya\textsuperscript{119} carried out a similar experiment to Busschbach and co-workers.\textsuperscript{99} Younger respondents generated the following weights for the ages 5, 10, 35, 60 and 70 years: 1.8, 1.6, 1.0, 0.5 and 0.6, respectively. The weights from older respondents were 0.6, 0.8, 1.0, 0.5 and 0.3, respectively.
Suggested equity weights
Weights for personal value and productivity
The first two views outlined above have been described as supporting ‘efficiency-based’, rather than ‘equity-based’, age weighting.\(^{115}\) This is because the concern is with the total sum of value produced by healthcare, rather than the impact of healthcare on the distribution of health. This value might be based on an individual’s assessment of the value of being healthy at different ages in terms of personal benefits to be derived from health. It might also be based on the increased value to society derived from the health of its most productive members. Setting priorities for treatment on the basis of a relationship between a person’s productivity and age, has been called productivity ageism.\(^{115}\)

Do people hold ‘efficiency-based’ preferences in relation to age?
The studies by Williams\(^{114}\) and Busschbach and co-workers\(^{99}\) provide the most direct evidence for these types of preferences. Williams asked people to consider specifically the value of health at different stages in life. In this way the study tried to focus the respondents on the value of health and to avoid distributional judgements. Busschbach and co-workers\(^{99}\) designed an experiment which would elicit specifically the types of preferences related to personal and social value. This was achieved by equalising the amount of lifetime health experienced by each hypothetical person in the experiment, to remove equity considerations. A pattern of weights was found, in which the value of health increased through childhood, hit a peak at some point and declined into middle and older age. Exactly where it might peak was not determined. The preference of this sample for health at age 35 years over health at age 60 years is consistent with productivity ageism. However, the health of a 10 year old was preferred to that of a 35 year old, indicating that perhaps the study subjects judged that health is of greater benefit to a person at age 10 years than at age 35 years. Considerations of productivity were thus not the only factor brought to bear on choices.

Further evidence from the studies reviewed supports both these findings. Lewis and Charney\(^{100}\) also found that the health of older children was preferred to the health of very young children. Cropper and co-workers\(^{116}\) found evidence consistent with discrimination on the basis of productivity: in their study, the health of a 30 year old was given more value than that of a 20 year old.

The general trend of the empirical results (a preference that younger people should be prioritised over older people) is consistent with a view of equity as a concern for equality in lifetime health. However, a few studies are inconsistent with this trend, as already noted. Cropper and co-workers,\(^{116}\) Busschbach and co-workers\(^{99}\) and Lewis and...
Charny found that, in the comparison of certain age groups, the older person is preferred to the younger. This indicates that age-related preferences are complex.

A difficulty in interpreting the results of these studies is that in making choices between saving different numbers of people at different ages, study subjects were free to express attitudes to more than one principle. For example, Cropper and co-workers and Johannesson and Johansson allowed health gains to vary between the age groups in their experiments. This, in principle, allowed their respondents to bring judgements of efficiency (health maximisation) into their responses. In addition to this, they would also have been able to express their views on the benefits of health to individuals at different ages and to productivity at different ages. Nord and co-workers, however, did control for differences in health gain. But judgements about productivity and its relationship to a person’s age were not controlled for. Thus, it is conceivable that the resulting weights are a composite of three factors: health maximisation, productivity and equity.

Alternative methods for incorporating equity concerns in economic evaluation

The willingness-to-pay approach
The framework of cost–benefit analysis, involving the measurement of health benefits in utility terms by means, for example, of a willingness-to-pay approach may represent an alternative for incorporating equity concerns into economic evaluation. If the analyst can address the issue of how the initial distribution of welfare and ability to pay may influence the resulting prescriptions of cost–benefit analysis (an issue that would require extensive discussion, beyond the scope of this project), cost–benefit analysis may offer the advantage of providing a more suitable framework than cost-effectiveness analysis for assessing interdependent utilities. However, this would not eliminate the need to elicit societal values and equity weights to support the formulation of a social-welfare function. The two aspects are complements rather than substitutes. The combined measurement of individual interdependent utilities and societal values for alternative distributions would pose new conceptual and empirical problems. As pointed out by Labelle and Hurley, there is a potential risk for double-counting due to the close relationship between individual and societal values, but there may also be conflicts between the two. If the view that efficiency-based distributions (driven by interdependent utilities) are impossible and undesirable, as argued by Mishan, is accepted, the cost–benefit analysis framework and the willingness-to-pay approach to measuring the benefits of healthcare seems to offer limited advantages over cost-effectiveness analysis. In any case, both researchers and users of economic evaluations seem to have favoured the cost-effectiveness approach in the last three decades, and therefore the discussion here of the willingness-to-pay solution as a possible way forward in the assessment of the distributional implications of allocative choices in healthcare will be limited to these remarks. It is unlikely that a claim of superiority of cost–benefit analysis compared to cost-effectiveness analysis, on the grounds of a greater ability to incorporate distributional judgements, would change the established trends in the use of this approach in healthcare economic evaluation.

The person trade-off method and cost–value analysis
Nord has argued that the person trade-off technique (PTO) offers a possible alternative to equity weighted cost-effectiveness ratios as a method for incorporating fairness into the analysis of healthcare interventions. In Nord’s view, the PTO has properties that make it the most suitable technique for valuing health outcomes. He suggests that PTO-based evaluation should be called ‘cost–value analysis’. Below we briefly describe the PTO method and how it might incorporate equity.

The PTO is primarily a method for eliciting values. However, two further features have been ascribed to this technique:

• as offering a method that elicits ‘social’ as opposed to individual (or private) values from respondents
• as offering a method that allows the respondent to incorporate more than one relevant aspect of a resource allocation problem, into a single value.

Below we describe the PTO technique for deriving values. We then discuss the view of Nord that the PTO could form the basis for a new kind of analysis.

The PTO as a method for eliciting values
The PTO can be used as a method for evaluating the value of a single characteristic such as health gain or age. In broad terms, the technique involves...
asking study respondents to balance up numbers of people who differ with respect to some given characteristic by asking how many people of one kind are equivalent to a given number of individuals of another kind. We have seen earlier in this chapter (see page 26) that the technique has been used to derive age weights. Respondents were asked questions such as: How many lives of people aged 50 need to be saved to be equivalent to 10 lives saved of people aged 20? When using the PTO technique to value health states the researcher specifies a reference group of individuals who are returned to full health from a reference health state (such as a given disease). An intermediate health state can then be valued by asking how many people of one kind have to be returned to full health from this intermediate state to be regarded as equivalent to the health gain in the other members of the reference group. The value of the intermediate health state is then calculated as the ratio \( \frac{x}{y} \). In answering PTO questions of this type, respondents perform the task of balancing up the number of people who benefit against the total health gain (e.g. Nord and co-workers87, Ubel and co-workers88).

PTO and cost–value analysis

Nord120 comments: “The person trade-off technique is a way of estimating the social values of different healthcare interventions” (our italics). Olsen122 also suggests that the PTO has this property: “The first method is a person–trade-off type which implies that social weights are attached to years with improved health when these are distributed between groups of people” (our italics).

The claim made for the PTO is that the study respondent gives a different type of answer to that elicited from techniques such as the standard gamble or the time trade-off. When answering the PTO question, the respondent is required to make a judgement between groups of people with particular characteristics, rather than express their own valuation from the perspective of what a particular health status would mean to them. Nord argues that health-state values based on these judgements have greater validity for use in social resource allocation decisions.

In his book, Nord120 states: “… the person trade-off approach allows encapsulation of concerns for fairness” (page 117). Which concerns for fairness? Nord discusses two particular aspects of equity that in his view could be incorporated in analysis via the PTO method. These are concern for severity in pretreatment health status, and the concern that cost-effectiveness analysis places too much emphasis on the role of health gain in a fair allocation of resources. But what about other aspects, such as age or social inequalities in health? It is not clear how these are to be built in. In fact, it is unclear in general whether Nord envisages PTO weights attached to years of life gained, in which case cost–value analysis would seem to be cost-effectiveness analysis under a different name, or whether he envisages the replacement altogether of years of life as part of the metric of health valuation. In principle, the PTO technique could be used to value the outcomes of healthcare interventions, incorporating equity and efficiency concerns in a composite measure. One could select a reference group of individuals with particular characteristics (e.g. a particular age, sex, ethnicity, disease, prognosis without treatment, prognosis with treatment, overall life-expectancy) for comparison via PTO to derive the ‘social value’ of healthcare interventions. This would, in principle, produce an overall index incorporating all factors of relevance. But Nord does not explicitly describe such a procedure.

The descriptive approach: presenting tabulations of effects

An approach that is significantly different from those examined so far involves giving up the aspiration to address equity concerns normatively within economic evaluation and pursuing a positive approach that would leave to decision-makers the responsibility to make judgements about the desirability of alternative distributions, on the basis of their own values or their perception of societal values. However, it must be recognised that economic evaluation always incorporates, to some extent, normative statements, and that the results presented, even in a descriptive framework, would not be neutral from a distributional viewpoint, as illustrated earlier in this chapter (see pages 14 and 15). What makes the tabulation of effects a positive solution, in the sense adopted in this report, is the fact that the analyst does not introduce his own norms in assessing the value of effects occurring to different population groups and in combining such effects.

A positive approach to addressing equity concerns would involve providing decision-makers with all the information they need to make a judgement on the desirability of alternative distributions. When cost-effectiveness analysis is applied on the basis of the decision rules described earlier in this chapter (see pages 15 and 16; which imply, among other things, that a uniform outcome measure is used), the key distributional effects produced by resource allocation decisions are those determined by the fact that different populations may benefit from
different interventions. In order to enable decision-makers to apply their own values about equity in deciding what interventions should be provided, the following pieces of information should be made available to them:

- information on the characteristics of the population that may potentially benefit from each intervention, along a number of dimensions relevant to equity that could be uniformly predetermined (e.g. age profile, gender mix, ethnic composition, socio-economic profile)
- information on how the effects of interventions (costs and effectiveness) may vary in different subgroups of the population that may benefit from such interventions, again along a number of dimensions that may be predefined for all economic evaluations
- when interventions could, in principle and in practice, be provided selectively to certain subgroups, information on the cost-effectiveness of providing those interventions in each subgroup should be supplied – the way in which these cost-effectiveness ratios should be calculated varies according to whether provision to different subgroups would represent a set of mutually exclusive interventions (e.g. when an intervention is, by its nature, in short supply) or not – in the former case incremental ratios should be determined through direct comparisons between subgroups, as suggested by the decision rules illustrated earlier in this chapter (see pages 15 and 16) for clusters of mutually exclusive interventions, while in the latter case there should be only one comparator for all options.

Discussion

As discussed at length in this chapter, preferences over alternative distributions originate partly from individual interdependent utility (mainly related to altruism) and partly from societal ethics (or the value system upon which the social contract is based). Economic evaluation has been developed as a normative tool to aid resource allocation decisions, and should ideally take both aspects into consideration if it is to address equity concerns. However, the ability of economic evaluation to retain its normative strength depends crucially on an accurate reflection of individual and societal values in the analysis. Attaining this goal poses formidable conceptual and methodological difficulties. In the first place, the conjoint measurement of individual and societal values is difficult due to the complex relationship that exists between these two sources of preferences over alternative distributions. Interdependent utilities partly generate societal preferences, leading to a significant potential risk of double-counting, but in some cases these may be in conflict with societal ethics, as pointed out by Mishan. This problem has so far been avoided by ignoring interdependent utilities. The debate about the nature and the measurement of this form of externality has been virtually non-existent, with the exception of a study published by Labelle and Hurley, which constitutes a useful but only initial step forward. Perhaps the reason why interdependent utilities have received less attention than they deserve is that these cannot be meaningfully combined with the non-preference-based outcome measures widely used in cost-effectiveness analysis.

Attempts to address equity concerns through economic evaluation in a normative framework have mainly relied on the elicitation of values that may support the definition of a social-welfare function. An apparent contradiction is that many empirical studies have tried to elicit societal values by conducting surveys of individual preferences. The latter may arguably reflect interdependent utilities more than societal ethics, although the adoption of the ‘veil of ignorance’ approach might overcome possible conflicts between the two perspectives. The values elicited may constitute a basis for determining a set of equity weights, reflecting a social-welfare function, to be used as adjustment factors for health outcomes in cost-effectiveness and cost–benefit analysis.

The review illustrated in this part of the report provides answers to three key questions regarding the prospects for equity weighting as a solution to the problem of incorporating equity into economic evaluation:

- What do we currently know about individual/societal preferences in relation to equity in healthcare?
- Is this knowledge sufficient to derive an adequate system of weights?
- What methods should be used to derive these weights?

What do we currently know about individual/societal preferences in relation to equity in healthcare?

Empirical studies were reviewed in three parts:

- empirical investigations of the equity–efficiency trade-off
• studies of views towards characteristics such as social class, sex or ethnicity
• studies of views towards age.

The results of work relevant to the equity–efficiency trade-off generally revealed a rejection of health maximisation as the sole criterion for allocating resources. This is consistent with what would be expected. The evidence revealed four types of trade-off:

• between health gain and the number of people treated
• between the total amount of future health in a population and the degree of equality between groups
• between the total amount of health (past and future) in a population and the degree of equality between groups
• between severity of illness and health gain.

These studies provide indications that people may make such trade-offs, but do not provide useable weights. There are several reasons for this, including the influence of framing effects on results. This, and other methodological issues, are discussed further below.

With regard to equity between social groups such as occupational class, sex and ethnicity, existing surveys were generally not designed to ascertain whether inequalities that might endure between these groups matter to people. We therefore lack information about whether inequalities between such groups would be regarded by people as an appropriate basis for setting priorities. Two studies did examine such an issue.97,103 However, these were studies of the views of politicians in Sweden, and therefore may lack relevance to the UK.

Age was the most intensively researched equity dimension. Despite this, it is not possible to infer a particular set of age-related equity weights from existing studies. As discussed in the section on weights for equity, the design of these studies was not adequate to reveal the weight given to particular principles, making interpretation difficult. Thus, although there are a few studies that report sets of age weights, these are not suitable for use in cost-effectiveness analysis, since it is not possible to determine precisely which equity principle they represent. Furthermore, although a general pattern of preference for younger over older people would seem to exist, this is not found across the entire age spectrum. The empirical work suggests that any ‘solution’ to the role of age will be complex.

Our review has shown that the current evidence base is small. Nine quantitative studies (i.e. studies which derived weights) were found, of which eight were English-language studies87,91,95–97,99,116,117 and one was Japanese.119 Samples were typically small and unrepresentative, often relying on convenience samples of students (as was the case, for example, in Andersson and Lyttkens,96 Johannesson and Gerdtham95 and Dolan91). However, there are two methodological issues of even greater concern in the assessment of preferences over alternative distributions and related equity weights: framing effects and unidimensional assessment. These issues are important both in interpreting existing evidence and as challenges facing any future research.

Is our knowledge sufficient to derive an adequate system of weights?

Our review of empirical studies of public opinion towards rationing healthcare shows that current knowledge of societal preferences in the UK is not sufficient to form the basis for a set of equity weights. This is particularly the case in relation to quantitative assessment of the relative weight that people place on health outcomes based on trade-offs between principles of efficiency and equity. At the time of writing, only one quantitative study to derive weights has investigated preferences in a UK sample.91 We noted in the section on sources of equity weights that a possible alternative approach to equity weighting involves the assignment of weights by the analyst based on patterns observed in public opinion surveys. However, to serve as a basis for equity weights, such surveys need to target people’s reasoning in quite a specific way. Surveys which have asked respondents to rank a selection of programmes will tend to lack the precision to allow inference of why people have made particular choices, and the relative weight they have given to different factors. It may be difficult to infer a set of weights from such surveys. An alternative survey technique has been to ask whether particular groups of people should receive higher or lower priority. This again leaves unanswered the question of the degree of priority and whether there are any trade-offs. The issue also remains that to introduce weights for a particular concept of equity requires understanding the reasoning of survey respondents.

What methods should be used to derive these weights?

Interpretation of the results of the empirical studies reviewed in this report should take account of the influence of the framing of questions. A
number of framing effects are likely to influence responses, including the precise context in which questions are set, the design of questions and, in the case of the quantitative studies, the methods for eliciting the strength of preferences.

An important aspect of the influence of framing relates to the problems of bias encountered in designing questions. For example, the relative weights derived in the age-weight studies might be sensitive to the particular ages chosen for comparison. Cropper and co-workers found that, where the age of 30 years was valued as 1, 60 years of age was weighted at 0.1; in the study by Nord and co-workers, age 20 was weighted as 1 and age 60 years was weighted at 0.4. This difference between the two might be due, in part, to framing effects. Framing effects of various kinds are widely recognised as a methodological difficulty in survey design. They have been shown to influence responses to quality of life questionnaires and can relate to how health outcomes are presented (e.g. responses to quality of life questionnaires and can design. They have been shown to influence effects. Framing effects of various kinds are widely recognised as a methodological difficulty in survey design. They have been shown to influence responses to quality of life questionnaires and can relate to how health outcomes are presented (e.g. responses to quality of life questionnaires and can relate to how health outcomes are presented (e.g. Cropper and co-workers found that a question was framed in terms of death rates or survival systematically influenced responses), how language is used (Gerard and Dobson found a significant difference in values depending on whether or not the word ‘cancer’ was used) or the order in which people have been asked to consider different options in eliciting preferences. It is plausible that similar types of problems will beset studies of preferences towards equity. The small literature on public preferences towards equity has yet to address these issues.

In relation to equity weights, the perspective that study respondents are required to adopt in making their judgements might affect the answers they give. Two empirical studies reviewed in this chapter utilised a device known as the ‘veil of ignorance’. This approach differs from other techniques in that the person taking part in the study is asked to imagine that they are one of the affected people in the decision problem (either a fortunate person with higher life-expectancy, or a less fortunate person with a lower life-expectancy) but without knowing which. This perspective may generate a different set of weights from one in which the person making the judgement is, within the hypothetical context of the experiment, excluded from the consequences of their decision. Nord has also argued that perspective is important to the answers people will give, proposing that the PTO offers a perspective that is social, rather than private (and therefore, in his view, appropriate for the derivation of equity weights) (see pages 28 and 29).

A further framing issue is that which relates to the influence on study results of the metric in terms of which respondents are asked to express trade-offs. As the review showed, some studies have used the person trade-off, particularly the studies of age. Others used health (e.g. studies of the equity–efficiency trade-offs). The choice of metric has been shown to influence choices in experiments to derive quality-of-life weights, and thus may be expected to do so in the case of equity weights.

In terms of the context in which questions are set, there may be a multitude of factors which influence answers to questions. For example, the following might be expected to have a bearing on what people think is fair: the way in which the disease(s) in question affect quality of life; whether, or the extent to which, life-expectancy is affected; the way in which the disease or illness is acquired; expected quality and length of life following treatment; the age of the individuals or groups of people concerned; the past health of the individuals or people concerned; past and expected future general welfare of the individuals or groups of people concerned; the overall health of relevant social groups in society to which the individuals in the rationing problem belong (e.g. occupational classes); the contribution to society of the individuals in the rationing problem with and without treatment; and the past contribution to society of the individuals in the rationing problem.

The contextual factors listed above, and maybe others, could influence the direction of people’s preferences and the weights derived from their choices. We currently lack information on how different combinations of these factors affect responses. For example, how would age interact with different degrees of severity? Would it be meaningful to derive two sets of weights separately, one set for age and one set for severity, and then add them together? To do so, would be to discount the possibility of a more complex relationship. Clearly the picture would become even more complicated if three or four dimensions were included.

The empirical work reviewed in this report tended to be unidimensional: the studies allowed only one factor to vary (exceptions being the study by Nord and co-workers, which derived age weights for both life-saving and morbidity-reducing interventions separately, and the studies by Johannesson and Gerdtham and Andersson and Lyttkens which allowed the initial inequality between social groups to vary). The results reported by Andersson
representative sample of the population, since they dynamics on post-discussion views). They may also problems of bias (e.g. the impact of group However, these methods may have their own effects. Methods that offer more time for discus-

tions remain over which dimensions of equity should be included. Secondly, if weights obtained for one particular feature (e.g. age) are found to be dependent on values of other factors, any effort to derive a meaningful set of equity weights will need to consider a large number of rationing decision contexts.

Public opinion might be consulted in a number of different ways, such as through surveys, focus groups or citizens’ juries. There are advantages and disadvantages associated with each of these methods. Surveys offer the potential for obtaining responses from a large, representative sample of the general population. However, this method may be particularly vulnerable to artificial framing effects. Methods that offer more time for discussion and reflection, such as focus groups or citizens’ juries, may offer a way of ameliorating these sorts of framing effects. It has also been suggested they will produce responses that are better reflections of what people really think (indeed, there is some evidence that the experience of a focus group changes people’s views\textsuperscript{108}). However, these methods may have their own problems of bias (e.g. the impact of group dynamics on post-discussion views). They may also tend to be based on a more highly selected, less representative sample of the population, since they require a greater degree of commitment on the part of participants than that required in completing a survey.

Conclusions

The conceptual and methodological issues discussed herein prompt the question of whether a normative approach to addressing equity concerns through economic evaluation is a viable option. Adopting the framework of cost–benefit analysis, which has proven less attractive in the healthcare domain, would solve only some of the problems highlighted. The obvious alternative to a normative approach is a positive framework within which information on the cost-effectiveness of health interventions would be provided alongside information on the distributional effects of such interventions, and the task of making an overall judgement about the desirability of alternative allocations would be entirely left to decision-makers. This approach may be based on the tabulation of the effects of health interventions along prespecified equity dimensions. The information that should be provided by economic evaluations within such a framework depends on the type of equity question that is to be addressed. For instance, all evaluations reporting incremental cost-effectiveness ratios may potentially be used to support resource allocation decisions involving the prioritisation of interventions competing for a given pool of resources. In order to address this equity question, economic evaluations should describe in detail the characteristics of the populations that may benefit from alternative interventions, and should illustrate the effects of such interventions in different subgroups along any relevant equity dimensions. When interventions could be provided selectively to subgroups, provision to each subgroup should be considered as a separate intervention and incremental cost-effectiveness ratios should be determined accordingly.

We present in Table 1 an assessment of the methods of tabulation and equity weighting against three criteria: feasibility, credibility and consistency. ‘Feasibility’ refers to the methodological challenges and costs involved in producing valid information for use in addressing equity issues within economic evaluation. The methodological challenge is whether valid, meaningful estimates of equity weights can be derived from preference elicitation studies. The costs may be prohibitive in some cases, but not in others (of course, whether the costs are prohibitive will be a judgement that has to be made
on a case-by-case basis). ‘Credibility’ refers to whether the information would be regarded as meaningful and as providing a legitimate basis for decision-making, by policy-makers and perhaps also the public. Credibility is therefore about the political acceptability of the information provided by economic evaluations as a basis for equity judgements by policy-makers. ‘Consistency’ refers to the question of whether using the information provided by these approaches would lead to a consistent pattern of decisions by decision-makers. Will the information provided by economic evaluation, to the extent it is used in a decision-making process, tend to produce a consistent set of decisions? For example, will the (implicit or explicit) weights given to age tend to be the same or vary across decisions?

### Feasibility

Tabulation faces the challenges of obtaining enough information to describe adequately the likely distributional effects of an intervention (effects and costs in subgroups, the characteristics of the benefiting population). This information may not always be available and obtaining it may be costly (e.g. the costs of running larger clinical trials). However, the application of equity weights requires both this information plus a set of valid weights. We have placed a question mark over whether it is technically possible to produce a valid set of equity weights. This is in recognition of the considerable methodological challenges of doing so. However, this does not mean that these challenges cannot be overcome.

### Credibility

Tabulation is likely to be seen as possessing transparency since its aim is to provide a full description of the distributive implications of an intervention. The information provided should be interpretable by users. As such it could be anticipated to possess credibility. However, a question mark hangs over the likely credibility of equity-weighted cost-effectiveness ratios. Potential users will have to trust in the validity and meaning of the numbers in front of them. Ultimately it might be necessary that those affected by decisions potentially based on equity-adjusted cost-effectiveness ratios, also trust and accept these calculations.

### Consistency

We have suggested in Table 1 that decisions made on the basis of descriptive information might in some cases lack consistency. It is possible that decision-makers would bring different considerations to bear, or may weight different dimensions of a problem differently, on different occasions. Also, decision-makers may change and bring with them different views. We have therefore placed a question mark over whether consistent decision-making will occur through the use of tabulated effects. Equity weights, in principle, should bring some consistency to decisions through stabilising the weight given to various dimensions relevant to equity. However, consistency would only be ensured by a systematic compliance of decision-makers to the prescriptions of economic analysis, which is a condition that could be attained by normative analysis only in ideal circumstances.

---

**TABLE 1 An assessment of tabulation and equity-weighted cost-effectiveness ratios as alternative methods of addressing equity through economic evaluation**

<table>
<thead>
<tr>
<th></th>
<th>Feasibility</th>
<th>Credibility of information and decisions</th>
<th>Consistency in decision-making</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tabulation</td>
<td>✓</td>
<td>✓</td>
<td>?</td>
</tr>
<tr>
<td>Equity-weighted cost-effectiveness ratios</td>
<td>?</td>
<td>?</td>
<td>✓</td>
</tr>
</tbody>
</table>

✓, performs well; ?, uncertain performance
Chapter 3

A review of published economic evaluations

Research questions

In this section we report the results of a review of economic evaluations published in five sample years (1987, 1992, 1995, 1996, 1997) aimed at determining the state of the art of the economic evaluation methodology with regard to its ability to address distributional questions. In particular, the following research questions have been investigated:

• What is the size and what are the characteristics of the literature addressing healthcare resource allocation problems that may have significant distributional effects?
• Are distributional effects explicitly incorporated in cost-effectiveness ratios or cost–benefit indexes in published economic evaluations, using one of the methods discussed in chapter 2?
• When not explicitly incorporated in calculations, have distributional effects been considered and has sufficient evidence been provided to decision-makers to enable them to make a judgement on the desirability of alternative allocations?
• Overall, do existing economic evaluations provide a satisfactory guide for addressing both efficiency and equity concerns involved in resource allocation decisions?

As a side issue, but no less important than those listed above, we explored and assessed methods for conducting systematic reviews of economic evaluations. Given the lack of established methods and the difficulties involved in searching the literature for economic evaluations, we pursued the additional aim of devising a sound search strategy that can be adopted within future reviews.

Systematic review methods

Time period and literature sources

The review was conducted on five sample years: 1997, 1996, 1995, 1992 and 1987. These include the three most recent years for which relatively complete records of published literature were available in the main electronic databases in October 1998, when the search process was started, and two sample years in the earlier period, five and ten years before the latest year available.

In developing our literature search strategy, we considered a number of widely used electronic databases covering the medical and social science literature: Index Medicus/MEDLINE, Excerpta Medica/EMBASE, HealthSTAR, PsycLIT, EconLit, Public Affairs Information System (PAIS) and IBSS. These are briefly described in appendix 1.

Definition of economic evaluation and other inclusion criteria

The selection procedure adopted in the literature review involves a sequential screening of the references identified by the search strategy described in the following sections. This procedure is also illustrated graphically as a flow chart in appendix 2.

The search was limited to studies assessing healthcare interventions directly aimed at improving the health of individuals or populations. The selection of references identified by the search strategy was carried out in two stages. Initially, references were screened on the basis of their titles and abstracts, excluding those that did not have an abstract and those for which there was sufficient and unequivocal evidence that inclusion criteria would not be met. Full articles were subsequently retrieved for the remaining references and checked against inclusion criteria.

In a test search described in the following sections, using the search strategy adopted in this review, the first stage of the selection led to the exclusion of approximately three-quarters of the references (73%). A further 9% of the original sample was excluded in the second stage. The remaining studies (18%) were classified as economic evaluations.

The definition of economic evaluation adopted in the selection is based on the approach proposed by Drummond and co-workers. According to this, "two features characterize economic analysis, regardless of the activities (including health services) to which it is applied. … First, it deals with both the inputs and outputs, sometimes called costs and consequences, of activities. … Second,
economic analysis concerns itself with choices”. Therefore, economic evaluation is defined as “the comparative analysis of alternative courses of action in terms of both their costs and consequences”. Accordingly, abstracts and, when appropriate, full papers were checked for evidence of the measurement of both costs and outcomes of alternative programmes.

Review articles that did not present new analyses were excluded. Meta-analyses or modelling studies based on data from meta-analyses were included. Cost analyses were excluded, whereas cost-minimisation analyses (“when the consequences of two or more alternatives are examined alongside costs, and are shown to be equivalent”) were considered as cost-effectiveness analyses indicating dominance of one alternative.

All outcome measures were considered acceptable, including intermediate or surrogate measures, as long as evidence was available of the benefits of achieving such intermediate outcomes. Studies with multiple outcomes that did not indicate a primary outcome as a possible basis for a cost-effectiveness ratio (cost–consequence analyses) were considered as cost-effectiveness analyses only when revealing a situation of clear dominance of one of the alternatives.

Studies comparing one intervention to a ‘do nothing’ alternative were included only when this was explicitly stated and an appropriate marginal analysis was undertaken.

Finally, we set additional inclusion criteria for the specific purposes of this review, which is aimed at assessing how economic evaluations address distributional effects. In fact, some resource allocation decisions may have no distributional implications, depending on the nature of the interventions, their cost and outcome profile. In particular, whenever a comparison of two alternative interventions for the same condition reveals a situation of dominance of one intervention (lower cost and higher effectiveness), the resulting choice (supposedly in favour of the dominant option) is unlikely to have any effects on equity grounds. This is generally true, except in two types of comparison: those in which the patient groups that may benefit from the two interventions are different (e.g. medical versus surgical therapy); those in which the two interventions involve a different distribution of costs between patients and other payers. However, these two cases are relatively rare, and therefore studies indicating dominance of one intervention were considered irrelevant and were excluded from the review.

Search strategies
Two extensive bibliographies of economic evaluations compiled by Elixhauser and co-workers indicate that the majority of such studies are published in medical journals, especially in more recent years. Therefore, the primary focus of our attention was the main medical literature indexing system (Index Medicus/MEDLINE). Other medical and social science databases were searched only incrementally (i.e. excluding references already contained in MEDLINE, regardless of whether these would be identified by our search strategy for the latter). Given that most evaluations were identified through MEDLINE, as indicated later in this chapter, a thorough test of the performance of alternative search strategies was undertaken only for this database.

The main issue in searching medical literature databases for economic evaluations is finding an acceptable balance between the sensitivity and the specificity of the search strategy. In fact, due to the problems mentioned previously, particularly the inadequate indexing of studies, a very sensitive search strategy able to capture a proportion of economic evaluations close to 100% would necessarily have a very poor specificity and would retrieve an enormous number of references. This would pose a significant burden on the reviewer in terms of the analysis and selection of studies, and would make the review practically impossible without access to considerable resources. On the other hand, more selective strategies like those used in most published reviews of economic evaluations are likely to miss a significant number of relevant studies. In order to assess the performance of alternative search strategies and identify one with an ‘optimal’ balance between recall rate for economic evaluations and overall number of references to be screened, we carried out a test search on a limited period of time (January to March 1997). The search aimed at assessing the performance, in terms of recall rate for economic evaluations, of a number of selective search strategies (with relatively high specificity) against a comprehensive strategy considered a ‘gold standard’ for the identification of economic evaluations (very high sensitivity). As mentioned, the test search was conducted exclusively on MEDLINE. Alternative comprehensive and selective strategies considered are described in the following sections.
Comprehensive strategies for MEDLINE searches (electronic gold standard)

In order to identify a gold standard search strategy for the database MEDLINE with very high sensitivity in the selection of economic evaluations, we initially considered strategies already developed and available in the public domain. In particular, we identified two such strategies, one from the previously cited bibliography compiled by Elixhauser and co-workers, and the second from the Economic Evaluation Database of the UK NHS Centre for Reviews and Dissemination (CRD). In both cases the strategies were developed as methods for systematically searching MEDLINE for economic evaluations.

The strategy used by Elixhauser and co-workers is a combination of the following terms:

- Costs and cost analysis (Mesh)
- Cost–benefit analysis (Mesh)
- Cost (Text word)

Although ‘cost–benefits analysis’ is a MeSH subterm of ‘costs and cost analysis’ the authors found it convenient to undertake separate searches based on the two terms, starting with the former as this was more likely to include actual economic evaluations. Duplicates were identified and eliminated before screening the references retrieved by the broader terms less likely to contain evaluations. This strategy seems rather unspecific and retrieves a vast number of irrelevant references. In fact, the authors of the bibliography had to use very simple criteria for the selection of references, which led to the inclusion of many studies inappropriately labelled as cost–benefit or cost-effectiveness analyses. On the other hand, the strategy is also likely to miss a significant number of references by neglecting those classified solely by the subheading ‘economics’.

A second electronic search strategy that aims to be comprehensive is that used by the CRD of the University of York for the NHS economic evaluation database. This is based on the following terms:

1. economics.sh.
2. exp ‘costs and cost analysis’/
3. economic value of life.sh.
4. exp ‘economics, dental’/
5. exp ‘economics, hospital’/
6. exp ‘economics, medical’/
7. exp ‘economics, nursing’/
8. economics, pharmaceutical.sh.
9. exp ‘fees and charges’/
10. exp ‘budgets’/
11. (cost or costs or costed or costly or costing).ab,ti,kw,kp.
12. (economic$ or pharmacoeconomic$ or price$ or pricing).ab,ti,kw,kp.
13. or/1–12
15. editorial.pt.
16. historical article.pt.
17. 14 or 15 or 16
18. 15 not 17
19. ‘animal’/
20. ‘human’/
21. 19 not (19 and 20)
22. 18 not 21

Although very comprehensive, this strategy seems to have a far too large yield and low specificity for the individual researcher without access to the same level of resources available to the CRD. We analysed the yield of all the search terms used in the CRD strategy in the reference period (January to March 1997) and came to the conclusion that, while the first two are clearly key to the search, the other terms have a very small additional yield when searching for economic evaluations (using the definitions and criteria illustrated earlier in this chapter). Therefore, we finally decided to adopt a search strategy composed as follows:

1. ‘Costs and cost analysis’/
2. Economics.sh.
3. 1 and 2
4. ‘animal’/
5. ‘human’/
6. 4 not (4 and 5)
7. 3 not 6
9. (English or French or Italian or Spanish).la.
10. 7 and 8 and 9

This gold-standard strategy led to the identification of 5320 references for entire year 1997, only a few more than the 4951 that would have been identified by the strategy used by Elixhauser and co-workers, and many less than the 9446 that would have been identified by the CRD strategy (Figure 5). As the additional yield in terms of economic evaluations identified is small, it may be inefficient for an individual researcher to adopt the comprehensive strategy used by the CRD.

When applied to the reference period (January to March 1997), the strategy led to the identification of 1840 references. Authors’ names, bibliographical sources and all other information except titles and abstracts were deleted from the references retrieved and these were reviewed blindly by two
reviewers on the basis of the definitions and criteria illustrated earlier in this chapter. There were three reviewers in total. The 1840 references were evenly split into three groups, and each group of references was reviewed by a different pair of reviewers. Initial disagreements were resolved by direct discussion. Most of the references (1615 (88%)) were excluded at this stage because titles and abstracts provided sufficient evidence that these studies would not meet inclusion criteria. Full papers were retrieved for the remaining references. After analysing these (all but four that were unobtainable) against inclusion criteria, 96 additional references were excluded from the search, leaving a final set of 129 economic evaluations (7% of the original set of 1840).

At this point, alternative selective strategies (with higher specificity) were developed as illustrated in the following section, and were subsequently tested against the gold-standard strategy described herein.

Selective strategies for MEDLINE searches
We developed a number of alternative selective strategies with the aim of reaching an acceptable balance between recall rate for economic evaluations and manageability of the literature search, in terms of the number of references to be screened. For this purpose, we analysed carefully the titles, abstracts, MeSH terms and all other fields of the MEDLINE records corresponding to the 129 economic evaluations identified at the previous stage of the test search. This analysis revealed some clear patterns. The MeSH term ‘cost–benefit analysis’ (single term) appeared in the majority of the records included (83/129), whereas the MeSH term ‘costs and cost analysis’ (single term) appeared in 23 records (of which four also contained the subheading ‘economics’), and the subheading ‘economics’ appeared in 27 records. However, as ‘cost–benefit analysis’ is a subterm of ‘costs and cost analysis’, records including the former can also be retrieved by exploding the term ‘costs and cost analysis’. The terms ‘cost*’ and ‘cost-effective*’ (text word) appeared in a significant number of the records included, but the former also appeared in a significant number of those excluded. We were not able to identify any ‘journal’, ‘language’ or ‘institutional affiliation’ patterns, although English-language publications were clearly dominant. On the basis of this analysis we decided to test the following three selection strategies against the gold standard illustrated in the previous section:

**Strategy A**
1. ‘Costs and cost analysis’/
2. ‘animal’/
3. ‘human’/
4. 2 not (2 and 3)
5. 1 not 4
7. (English or French or Italian or Spanish).la.
8. 5 and 6 and 7

![FIGURE 5](image-url) The number of references retrieved using the electronic gold-standard search strategy and the other published comprehensive search strategies
Strategy B
1. ‘Cost–benefit analysis’/
2. ‘animal’/
3. ‘human’/
4. 2 not (2 and 3)
5. 1 not 4
7. (English or French or Italian or Spanish).la.
8. 5 and 6 and 7

Strategy C
1. ‘Cost–benefit analysis’/
2. exp ‘Costs and cost analysis’/
3. cost-effective*.ti,ab.
4. 1 or (2 and 3)
5. ‘animal’/
6. ‘human’/
7. 5 not (5 and 6)
8. 4 not 7
10. (English or French or Italian or Spanish).la.
11. 8 and 9 and 10

Strategy A was supposedly the most comprehensive (higher sensitivity), but we expected it to yield a very large number of references. On the other hand, strategy B was likely to yield a much more manageable set of references but was also likely to miss a significant number of economic evaluations. Strategy C was developed with the aim of capturing as many as possible of the economic evaluations missed by strategy B and identified by strategy A. Therefore we expected the yield of strategy C to be intermediate between those of strategies A and B.

The results of the analysis are reported in Figure 6, with reference to the whole of 1997. Strategy A identified 1131 references in the period from January to March 1997, whereas strategy B identified 469 references and strategy C identified 514. The latter strategy required the examination of approximately half the references retrieved by strategy A and less than 30% of the references retrieved by the gold-standard strategy, offering significant advantages in terms of manageability of the search. The performance of the three selective strategies in terms of recall rate for economic evaluations is illustrated in Figure 7. As expected, strategy A had the highest sensitivity, being able to identify 82% of the relevant studies, whereas strategies B and C led to the identification of 64% and 72% of the economic evaluations, respectively. Strategy A required the screening of 10.7 references to identify one economic evaluation, strategy B required 5.7 and strategy C required 5.5. The gold standard had a ratio of references to economic evaluations of 14.3.

Of course, there is no ‘optimal’ strategy. In the choice between alternative options researchers have to trade off sensitivity and specificity. Researchers aiming to undertake a systematic search may wish to use strategy A, or even the gold-standard strategy. If the search is limited to a set of conditions or interventions, researchers may be able to contain the overall number of references to be screened. On the other hand, when a more general search is to be undertaken, possibly on several years, strategy C seems to offer an excellent balance between recall rate for economic evaluations and overall number of references retrieved.
This conclusion was supported by our findings when evaluations were split into two subsets including studies indicating dominance and studies reporting incremental cost-effectiveness ratios, respectively. Out of the 129 economic evaluations selected, 52 reported incremental cost-effectiveness ratios, while 74 showed dominance of one alternative (one reported equivalence of the alternatives). When we re-assessed the performance of the three selective search strategies in the identification of economic evaluations reporting incremental cost-effectiveness ratios, strategy A identified 96% of these studies, strategy B 85% and strategy C 94%. This confirmed the excellent sensitivity of strategy C.

It must be said that the 129 economic evaluations identified in the search were a highly heterogeneous set. Although all satisfied the inclusion criteria, many studies were in fact rather trivial economic analyses aimed solely at demonstrating that one medical intervention was not only more effective than alternative interventions, but also less expensive. In many cases these studies reported very rough cost calculations and used intermediate or disease-specific outcome measures (aspects that were not assessed in the selection, because methodological quality was not a concern in this context). There is a significant overlap between studies indicating dominance of one alternative and those using less sophisticated methods.

**Strategies for searching other electronic databases**

Further searches were carried out on the electronic databases described later in this chapter. In most cases the subject headings were not as structured and detailed as in MEDLINE, and therefore the resulting search strategies were less accurate and retrieved a large number of non-relevant references. The performance of the strategies used was not systematically tested as was done for the MEDLINE search strategies, partly due to time constraints and partly because the yield of the additional searches was minimal compared to the yield of the MEDLINE search. The following search strategies were used:

**HealthSTAR**
1. ‘Cost–benefit analysis’/
2. exp ‘Costs and cost analysis’/
3. cost-effective*.ti,ab.
4. 1 or (2 and 3)
5. Journal article.pt.

**EMBASE**
1. exp ‘cost benefit analysis’
2. exp ‘cost effectiveness’
3. exp ‘cost’
4. 1 or 2 or 3
5. human
6. 4 and 5
7. abstracts
8. 6 and 7
9. english or french or spanish or italian
10. 8 and 9

**EconLit**
1. cost effect*
2. cost utility
3. cost benefit

---

*FIGURE 7 The number of references needed to be screened in order to identify one economic evaluation*
The searches of HealthSTAR, PsycLIT and BIDS (IBSS) did not provide any references not available in MEDLINE for the first three months of 1997. The search of EconLit provided one additional reference regarding a paper published in the Journal of Econometrics. However, a ‘clinical’ version of the same paper had been published previously in the journal JAMA, indexed in MEDLINE. The search of PAIS also led to the identification of one additional reference not present in MEDLINE. Finally, the search of EMBASE did not seem to add any references to the MEDLINE search in the period from January to March 1997. However, a further check of the yield of the search on several years revealed a dramatic and unexplained drop in the numbers of references retrieved by our search strategy in 1997. Given that the number of published economic evaluations has been increasing steadily, this may have been caused either by a change in the indexing system or by a significant delay in the inclusion of references in the database. To overcome this problem we ran the search for the entire year 1996 and compared the yield with that of a MEDLINE search on the same year. The search identified 25 references not present in MEDLINE that were potentially economic evaluations. The papers were not retrieved and assessed because we considered the number too small to justify adding EMBASE to the literature search. This was therefore conducted exclusively on MEDLINE, using the search strategy illustrated in previous sections, due to the negligible additional yield of other databases.

Analysis of economic evaluations
A checklist was devised for assessing the economic evaluations that met all inclusion criteria. This is reported in appendix 3. The checklist includes a number of items regarding general characteristics of the economic evaluations examined, and a number of more specific items regarding aspects of the evaluations that may help decision-makers address equity concerns.

In the general section, information on the year of publication, journal, author(s), the country where the study was undertaken, and the type of intervention examined was recorded. The characteristics of the design of the economic evaluations were examined in terms of technique used (cost–benefit versus cost-effectiveness analysis), perspective adopted, setting (or type of decision the evaluation was aimed to support, policy, managerial or clinical) and outcome measure (in the case of cost–benefit analysis it was determined whether a human capital approach or a willingness-to-pay approach was used).

The more specific section of the checklist is focused on the information provided within the papers about the characteristics of the population that may benefit from the intervention(s) and about the effects of the intervention(s) in that population or subgroups of it. First, it was determined whether the study examined a selected or unselected group of the population that may potentially benefit from the intervention(s). Secondly, the information provided on the characteristics of the population, or the selected group examined in the study, along dimensions that may be relevant for equity (e.g. age, gender, socioeconomic condition, ethnic group) was extracted from the papers, as well as information provided on how the effects of the intervention(s) may vary by the same characteristics. Other dimensions than those explicitly mentioned in the checklist were grouped in the category ‘risk profile’ (e.g. screening interval for screening programmes, subtype of disease, exposure to a particular risk
factor). Finally, it was determined how the study dealt with the distributional effects involved in the resource allocation decision (i.e. whether equity weighting (differential valuation of outcomes in different subgroups) was adopted, or whether a tabulation of unweighted cost-effectiveness ratios by subgroup was presented). A space was left at the end for a description of other solutions possibly used to deal with the distributional effects of the intervention(s) evaluated.

The studies were examined in terms of their ability to provide information on three key distributional effects:

- the effects of switching between the (mutually exclusive) interventions compared in an evaluation
- the effects of providing an intervention selectively to a subset of the overall population that may potentially benefit from that intervention
- the effects of prioritising between interventions competing for a given pool of resources.

Greater emphasis was placed on the last of the three questions, to which the inclusion criteria used in the literature search were more specifically targeted.

**Results of the systematic review**

The MEDLINE search strategy returned a total of 4951 references for the five years. Numbers increased over time, from 334 in 1987 to 1469 in 1997, as illustrated in Figure 8. At the end of the selection process, only less than 9% of these references were included in the review. The proportion of references included did not show any time-related pattern. This proportion varied between a high of 13% in 1992 and a low of 6% in 1995. In total, 424 studies were included (24 in 1987; 73 in 1992; 72 in 1995; 104 in 1996; 151 in 1997), as illustrated in Figure 8. The key characteristics of the studies are reported in appendix 5.

The countries of origin of the economic evaluations reviewed are presented in Table 2 by publication year. Overall, half of the economic evaluations (but as many as 88% in 1987) were from the USA. All other countries follow at great distance, led by the UK and Canada with 11% and 9% of the studies, respectively.

Studies are classified by intervention area and by year in Table 3. Economic evaluations of therapeutic interventions dominate, with 42% of the studies in the entire period, but if prevention and
screening are considered together, these amount to the same share of the total number of studies. Evaluations of diagnostic interventions (not including screening) are 10% of all evaluations. All but six of the evaluations reviewed are cost–benefit and a cost-effectiveness analysis. Three cost–benefit analyses were published in 1987 (two using the human capital approach, one based on willingness to pay), two were published in 1992 (one using the human capital approach, one based on willingness to pay), and two in 1997 (one using the human capital approach, one based on willingness to pay). Half of the cost-effectiveness analyses measure ‘final’ outcomes such as life-expectancy (20%) or quality-adjusted life-expectancy (30%), whereas the other half are based on ‘surrogate’ measures. The numbers and proportions of studies using different outcome measures are presented in Table 4 for each of the 5 years.

Economic evaluations could, in principle, address three types of distributional questions. These are related to the way in which resources are allocated in the healthcare system, and require different pieces of information from economic evaluations. The three questions are described below.

**Distributional effects of switching between the (mutually exclusive) interventions compared in the evaluation**

It could be argued that these distributional effects are relatively unimportant, and take place only when the patient groups that may benefit from the interventions compared within the evaluation are different, or when the interventions involve a different distribution of costs between patients and other payers. These distributional effects may arise regardless of the relative cost-effectiveness of the interventions compared (i.e. whether or not one

---

**TABLE 2** The number of economic evaluations included in the review, by year and by country

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>%</td>
<td>No.</td>
<td>%</td>
<td>No.</td>
<td>%</td>
</tr>
<tr>
<td>USA</td>
<td>21</td>
<td>88</td>
<td>30</td>
<td>41</td>
<td>36</td>
<td>50</td>
</tr>
<tr>
<td>UK</td>
<td>10</td>
<td>14</td>
<td>8</td>
<td>11</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>Canada</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Sweden</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Italy</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>6</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>France</td>
<td>7</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norway</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others*</td>
<td>3</td>
<td>14</td>
<td>16</td>
<td>16</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>100</td>
<td>73</td>
<td>100</td>
<td>72</td>
<td>100</td>
</tr>
</tbody>
</table>

*Includes countries with less than five studies in each column and studies on multiple countries

---

**TABLE 3** The number of economic evaluations included in the review, by year and by intervention type

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>%</td>
<td>No.</td>
<td>%</td>
<td>No.</td>
<td>%</td>
</tr>
<tr>
<td>Preventive</td>
<td>5</td>
<td>21</td>
<td>11</td>
<td>15</td>
<td>16</td>
<td>22</td>
</tr>
<tr>
<td>Curative</td>
<td>8</td>
<td>33</td>
<td>31</td>
<td>42</td>
<td>29</td>
<td>40</td>
</tr>
<tr>
<td>Diagnostic</td>
<td>5</td>
<td>21</td>
<td>7</td>
<td>10</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>Screening</td>
<td>4</td>
<td>17</td>
<td>22</td>
<td>30</td>
<td>15</td>
<td>21</td>
</tr>
<tr>
<td>Rehabilitation</td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Palliative</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>100</td>
<td>73</td>
<td>100</td>
<td>72</td>
<td>100</td>
</tr>
</tbody>
</table>
intervention is dominant). Examining whether distributional effects of this type may arise in a specific comparison is difficult, as it requires an in-depth knowledge of the clinical nature of the interventions compared and of the specific financing mechanisms in the setting where the evaluation was undertaken. This is beyond the scope of this review. However, in our examination of the economic evaluations included in this review (all studies in which no intervention was dominant) we were unable to find any studies addressing explicitly this type of potential distributional effects.

**Distributional effects of providing an intervention selectively to a subset of the overall population that may potentially benefit from that intervention**

These distributional effects are probably more important than those described above from a societal viewpoint, but arise only in particular circumstances. The most typical case in which interventions are provided selectively is when the funding for those interventions is capped on the basis of some rationing criterion. Examples of this include intensive care units, heart surgery and dialysis. In all these cases the interventions are offered by the health service, but the funding of these interventions is not sufficient to meet all the demand and waiting lists are formed. This leads inevitably to the prioritisation of patients for access to the intervention. Whether this is done on a ‘first come first served’ basis or on the basis of some more elaborate criterion, there will be distributional effects. A similar situation arises when constraints are determined by the scarcity of a key resource in the delivery of the intervention, rather than by insufficient funding. A typical example of this is the case of renal transplantation examined in chapter 4. In this case the scarcity of kidneys determines the need for prioritisation of patients, with important distributional effects.

Whenever conditions like those described are present, distributional effects arise regardless of whether the interventions in short supply dominate alternative interventions or not. In these cases economic evaluations should provide information on several aspects of the impact of the interventions examined in order to be able to support decisions led also by equity concerns. First, evaluations should focus on an unselected population of patients who may benefit from the interventions appraised, or at least on a sufficiently broad range of patient types, to allow comparisons between subgroups. Secondly, evaluations should explore the effects of the interventions (effectiveness and cost) along one or more dimensions considered important from an equity viewpoint. Finally, evaluations should report information on the relative cost-effectiveness of the interventions in different subgroups, or patient types, considering the delivery of such interventions to different subgroups of patients as independent (and mutually exclusive) interventions. In the process of determining the relative cost-effectiveness of interventions in different subgroups a differential valuation of outcomes (equity weighting), or another method for addressing equity explicitly, may be applied.

Again, the equity issues arising in the prioritisation of patients for access to a specific intervention was not the primary focus of this review because, in order to determine in what cases such issues potentially arise, a detailed exploration of the context in which an intervention is delivered would be required. However, none of the 424 evaluations included in the review reported the information needed to address these equity issues. Two thirds of

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality-adjusted life-expectancy</td>
<td>2</td>
<td>8</td>
<td>15</td>
<td>21</td>
<td>28</td>
<td>39</td>
</tr>
<tr>
<td>Life-expectancy</td>
<td>7</td>
<td>34</td>
<td>16</td>
<td>22</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>Survival rate</td>
<td>4</td>
<td>17</td>
<td>2</td>
<td>3</td>
<td>15</td>
<td>21</td>
</tr>
<tr>
<td>Cases successfully treated</td>
<td>2</td>
<td>8</td>
<td>10</td>
<td>14</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>Cases detected</td>
<td>1</td>
<td>13</td>
<td>16</td>
<td>22</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Cases prevented</td>
<td>4</td>
<td>17</td>
<td>12</td>
<td>16</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>Other</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>100</td>
<td>73</td>
<td>100</td>
<td>72</td>
<td>100</td>
</tr>
</tbody>
</table>
the studies focused on selected groups of patients. A relatively large number of studies did assess effects and determine cost-effectiveness ratios in subgroups, but none of the studies examined reported information on the relative cost-effectiveness of providing an intervention to one subgroup versus providing it to a different subgroup.

**Distributional effects of prioritising between interventions competing for a given pool of resources**

The distributional effects related to broad resource allocation decisions are probably the most significant ones and our review is primarily focused on these. Broad resource allocation decisions are typically made at the policy level and involve choosing among interventions aimed at treating different conditions and, consequently, patient groups that may differ in terms of age profile, gender mix, socio-economic conditions, ethnic group and other dimensions that are relevant in equity judgements. In this context, choices leading to the adoption of dominant interventions (those with better outcomes and lower cost than alternative interventions) are unlikely to have distributional effects, therefore the focus of this review is on interventions that produce better outcomes but require additional resources compared to alternative interventions that would be used if the former were not available. Distributional effects related to broad resource allocation decisions are also the primary concern of most theoretical contributions on equity and economic evaluation discussed in chapter 2.

For economic evaluations to be able to support resource allocation decisions driven by, among other things, equity concerns, a number of conditions must be met. First, evaluations must assess the cost-effectiveness of interventions based on outcome measures that allow comparisons across a broad range of interventions. It can be seen from Table 4 that a substantial number of studies focused on surrogate measures of outcome, which are unsuitable for broad comparisons. If we consider as final outcome measures quality-adjusted life-expectancy, life-expectancy and survival rates (or lives saved) – although it could be argued that the latter do not represent final outcomes – the proportion of potentially useful evaluations is only 64%. However, using a final outcome measure is a necessary but not sufficient condition. In order to be able to support decisions, evaluations should also provide information on the characteristics of the patient population that may benefit from the interventions appraised, and possibly provide a differential valuation of outcomes in different subgroups (equity weighting).

Only 35 of the 273 studies that compared interventions in terms of their final outcomes provide some explicit description of the characteristics of the populations that may benefit from such interventions. Twenty studies provide information on the age profile of the respective patient population, 21 provide information on the gender profile, and two provide information on socio-economic conditions. However, none of these studies presents comprehensive and structured information on patient characteristics, with only few and approximate details being provided. On the other hand, a relatively large number of studies (103) present some information on the effects of interventions in specific patient groups, and 71 of these present cost-effectiveness ratios by subgroup along dimensions such as age, gender and risk status, as illustrated in Table 5. There was no correlation between this characteristic and those illustrated in Tables 2 to 4 (country of origin, type of intervention and outcome measure). While at least some of these dimensions are certainly relevant in equity judgements, the utility of determining cost-effectiveness ratios in specific subgroups may be questionable, as discussed below. Differential valuation of outcomes in different patient groups

---

**Table 5** The number of economic evaluations adopting final outcome measures and reporting cost-effectiveness ratios, by patient subgroup characterised along different equity dimensions

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>3</td>
<td>6</td>
<td>16</td>
<td>10</td>
<td>20</td>
<td>55</td>
</tr>
<tr>
<td>Gender</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>19</td>
</tr>
<tr>
<td>Risk profile</td>
<td>1</td>
<td>4</td>
<td>8</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnic group</td>
<td></td>
<td>2</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
<td>6</td>
<td>20</td>
<td>16</td>
<td>25</td>
<td>71</td>
</tr>
</tbody>
</table>

*The number for different dimensions in each year do not add up to the total for the same year because studies may consider more than one dimension*
A review of published economic evaluations

(equity weighting) was not used in any of the economic evaluations included in this review.

Finally, none of the rare cost–benefit analyses assessing outcomes with the willingness-to-pay approach (only three in the five sample years), considers any utility aspects other than those related to the expected health effects of own consumption of healthcare in their willingness-to-pay assessment.

Discussion

The overall picture illustrated above is very disappointing, and certainly the worst that could have been expected before starting this literature review. Despite the significant and increasing number of economic evaluations published every year in the world literature, it can be said with confidence that none of those examined in this review, covering the three most recent years for which full records of published literature were available and two less recent sample years, provided enough information to allow decision-makers to judge the distributional consequences of alternative resource allocations.

The studies examined were unsuitable for addressing any of the three distributional questions described in the previous section. However, the selection of economic evaluations for this review was based on criteria that relate mainly to the third question (i.e. assessing the distributional implications of prioritising interventions in competition for a given pool of resources).

Only half of the economic evaluations reviewed measure outcomes that can be meaningfully used in comparisons of cost-effectiveness ratios across interventions (quality-adjusted life-expectancy, 30%; and life-expectancy, 20%); an additional 14% measure outcomes in terms of survival rates (or lives saved). Only about one in eight of these studies reported some information on the characteristics of the population that may benefit from the interventions appraised, thus shifting to decision-makers the burden of gathering such information from alternative sources. A larger number of studies reported information on the effects and cost-effectiveness of interventions in specific patient subgroups defined in terms of age, gender, risk profile or ethnic group. In practice, these studies evaluated the provision of an intervention to a specific patient group as an independent intervention. However, for this information to be useful in resource allocation decisions it should be possible to provide interventions selectively to specific patient groups, and this is extremely unlikely in most circumstances.

Cost-effectiveness ratios by patient subgroup may be more useful to address distributional questions of the second type illustrated in the previous section (i.e. for prioritising access to an intervention in short supply by different patient groups). However, in this case cost-effectiveness ratios should be calculated through direct comparisons between patient groups, rather than indirectly through comparisons of alternative interventions in different groups, as in all the evaluations reviewed.
Chapter 4

Distributional implications of recent health policies in the UK

Introduction

This part of the report includes three case studies of healthcare policies currently adopted in the UK: cervical cancer screening, renal transplantation and neonatal screening for sickle cell disease. The aim of these case studies is to show what conflicts and trade-offs may occur between efficiency and equity criteria in the implementation of actual health policies. The idea is to reveal the attitudes of NHS decision-makers in specific contexts, indicating what dimensions economic evaluations ought to take into account in order to provide relevant information in support of policy decisions.

The cost-effectiveness of each of the three policies is examined either on the basis of information from existing economic evaluations or on the basis of new evidence gathered as part of this project. Distributional implications (regarding socio-economic status in the first case study, age in the second, and ethnicity in the third), are also examined and implicit values with regard to equity are inferred, where possible.

The three case studies are not examples of good or bad policies, nor examples of good or bad evidence from economic evaluations. They are just meant to show the practical implications of neglecting distributional implications in cost-effectiveness analysis, and in what direction existing economic evaluation methods should be changed in order to provide more relevant evidence to decision-makers who face a trade-off between equity and efficiency objectives and wish to make an informed judgement about how this should be addressed.

The quantitative models used to assess the cost-effectiveness of cervical cancer screening and sickle cell screening are not illustrated in great detail in this report, because these are of marginal importance within the context of this report. However, details are available from the authors on request.

The national cervical cancer screening programme in England

Cervical cancer is the third most common cancer and a leading cause of cancer death for women worldwide. Age-adjusted incidence rates of invasive cancer of the cervix as high as 15–30 per 100,000 until the 1960s led many European countries to adopt organised national screening programmes. These have accomplished varying degrees of population coverage. In fact, mortality from cervical cancer has decreased dramatically over the last few decades in most European countries, although the issue of whether this may be mainly a consequence of underlying disease trends, rather than an effect of screening, is still controversial.

There is substantial evidence that screening for cervical cancer is both an effective and an efficient strategy (e.g. Quinn and co-workers, Eddy), but the size of the impact on mortality and healthcare resources is highly dependent on the context in which the programme is carried out and on the characteristics of the programme. In the UK, where mortality from cervical cancer has been particularly high, economic incentives to general practitioners (GPs) were introduced in 1990 to enhance the coverage and effectiveness of the national screening programme.

Differential and/or unequal availability and use of screening and treatment procedures across social groups could significantly affect the outcomes of screening. Both morbidity and mortality from cervical cancer have been strongly associated with socio-economic gradients. Therefore, reports that utilisation of health promotion and prevention measures is higher in the wealthy than in the less well-off should be of major concern in planning screening programmes. Previous research has highlighted that women most at risk of developing the disease, namely those from poorer backgrounds, were being screened less than others, both prior to and after the introduction of the organised screening programme in 1988.
The aim of this case study is to assess the cost-effectiveness and the distributional impact of cervical cancer screening in England and to examine possible equity-efficiency trade-offs. A simulation model has been developed for this purpose on the basis of existing information gathered from several sources. This was first employed to determine the cost-effectiveness of the current national screening programme compared with the strategy used before the introduction of economic incentives to GPs in 1990. Then, a simulation was carried out to assess whether the substantial increase in coverage obtained after the introduction of economic incentives to GPs (target payments) has been equally beneficial to women of all social groups and has contributed to reducing inequalities in morbidity and mortality from cervical cancer in England.

**Cervical cancer and health inequalities**

The evidence available of inequalities in cervical cancer morbidity and mortality is not merely anecdotal. Longitudinal studies undertaken by the UK Office of National Statistics have shown incidence rates varying from 20 per 100,000 in ‘manual class’ women to 12 per 100,000 in the ‘non-manual’ group (1976–89). The difference appears even larger in a later cohort, in which women belonging to the ‘manual’ group had an incidence rate of 22 per 100,000 (1986–89).137 Similar inequalities between women of different socio-economic conditions exist in survival after diagnosis of cervical cancer. Although these seem to be gradually decreasing, crude 5-year mortality rates for women diagnosed in the period 1986–90 range from 61% for the most affluent group to 56% for the most deprived.136

Given the multitude of factors which could possibly have a role in the development of cervical cancer, the task of establishing why it has a higher incidence in lower social classes becomes even more complicated and can only be hypothesised. Epidemiological research into different aspects of the aetiology of cervical cancer provides a number of possibly relevant findings:

- Cervical human papillomavirus infection, one of the causal factors of cervical cancer, seems to be more prevalent in women of lower education and income.142
- The widespread assumption that multiple sexual partners is the reason why cervical cancer occurs disproportionately among women from lower social classes has never been proven.143,144 In fact, empirical evidence seems to indicate that women from higher social classes have more sexual partners over their lifetime.145 Nevertheless, this does not seem to exclude the role of more promiscuous behaviour amongst lower social class men.143
- The age of first intercourse is earlier for women in lower social classes.145–146
- Women from lower social classes seem to use oral contraception more than barrier methods and to suffer disproportionately from sexually transmitted infections.145
- Higher tobacco use, an unhealthy diet and high levels of stress in lower socio-economic groups add up and have immunosuppressive effects which could enhance the persistence of the human papillomavirus.142

**Development of the national programme in England**

During the early 1960s individual district health authorities in the UK began to introduce cervical screening using the Papanicolaou (Pap) test. In 1967, the NHS announced a national programme which recommended screening at 5-year intervals for all women over the age of 35 years and set up a national recall system for women who had previously had a smear.136 Smears were delivered in general practice and GPs received a flat fee per woman aged over 35 years screened, once every 5 years. This was subsequently expanded to include women aged under 35 years, if they had had three pregnancies. Any other smears were not eligible for separate reimbursement and were paid for as part of the general per-capita fee a GP receives for each registered patient. A small number of women had (and still have) their smears taken at family planning or genitourinary clinics.146

At 31 March 1990 the screening programme covered approximately 64% of women in the age range 25–64 years at 5.5 years,146 or an estimated 59% of women aged 20–64 years at 5 years. In order to enhance coverage further, the
Department of Health introduced target payments, an innovative form of economic incentive to GPs, as part of the 1990 contract. The amount reimbursed to a GP for a smear was linked to the doctor’s ability to achieve coverage targets of 50% and 80%. The higher the coverage, the higher the payment per smear, with a fee ratio of approximately 3:1 between the two targets. Payments are ‘per woman’ rather than ‘per test’, which discourages GPs to make more tests than recommended. Women who have undergone a hysterectomy are excluded from target calculations. The relevant fee is paid annually, regardless of whether the GP took a smear during that year or not, as long as the woman had a test in the last 5 years (initially 5.5 years). The GP receives an adjusted payment if the woman had her test done in a community or hospital clinic; however, the woman will still be included in the target calculations.

The impact of the new incentives, combined with the computerised local call–recall system, was significant. Coverage at 5 years in England for women aged 20–64 years increased from 59% to 79% between 1989–90 and 1992–93, and reached 83% in 1997–98. Another study conducted in Denmark showed that the national guidelines for cervical screening issued by the National Board of Health in 1986 advocated a suboptimal strategy, with an incremental cost-effectiveness of approximately DKK 113,500 (assuming 80% coverage) compared to the next less effective alternative strategy. The authors examined a wide range of screening options and concluded that the lowest cost-effectiveness ratio could be obtained by screening women aged 30–59 every 5 years (the incremental cost-effectiveness versus targeting women aged 30–50 years was DKK 20,500, the average cost-effectiveness was DKK 21,000).

In the late 1970s an organised screening programme was progressively introduced in The Netherlands, initially aimed at women aged 35–55 years, to be screened at 3-year intervals, and subsequently extended to cover the age range 30–60 years, but with screening at 5-year intervals. In recent years the programme has achieved a national coverage of 75% up to the age of 50 years. After the introduction of the national screening programme in The Netherlands a reduction in mortality of 58% was achieved, although it must be said that rates fell by 42% between 1970 and 1980, when screening was very limited. Using a mathematical model it has been estimated that the current screening strategy may prevent 4563 deaths in the period 1993–2020 and save 118,824 life-years at a cost-effectiveness ratio (discounted at a 5% rate) of DFL 27,602 per life-year saved (equivalent to £9858 in 1993), compared with a hypothetical situation in which screening were not available.

Outside Europe, a national screening programme was implemented in Japan in 1961, although 30 years later coverage was still as low as 15.4%. The Netherlands started an organised screening programme in 1963 for women aged 30–55 years, with testing done every 5 years. In the early 1990s, following an increase of the upper age limit to 60 years, this invitation-based system achieved an overall coverage of 75%. It has been estimated that this programme reduced mortality from cervical cancer by 90%. Broadly similar programmes were introduced at approximately the same time in Sweden, Iceland and parts of Norway and Denmark. In these countries, however, the reduction of mortality from cervical cancer has been less evident. This is probably due to a narrower age range covered in Sweden (where age-standardised death rates for the age group 0–64 years fell by 78% in the period 1970–95, compared to 87% in Finland), and to a lower overall coverage in Norway and Denmark (where rates fell by 61% and 71%, respectively). Hristova and Hakama produced projections of the effectiveness, cost and cost-effectiveness of adopting the Finnish screening model in all Nordic countries. They concluded that the period 1998–2002 the programmes would have an incremental cost-effectiveness varying between US$1000 and 2600 per QALY gained for Finland and Denmark, respectively. Another study conducted in Denmark showed that the national guidelines for cervical screening issued by the National Board of Health in 1986 advocated a suboptimal strategy, with an incremental cost-effectiveness of approximately DKK 113,500 (assuming 80% coverage) compared to the next less effective alternative strategy. The authors examined a wide range of screening options and concluded that the lowest cost-effectiveness ratio could be obtained by screening women aged 30–59 every 5 years (the incremental cost-effectiveness versus targeting women aged 30–50 years was DKK 20,500, the average cost-effectiveness was DKK 21,000).

Evidence of the effectiveness and efficiency of screening

National screening programmes have been adopted in many European countries, particularly in Scandinavia and northern continental Europe. A national programme was started in Finland in 1963 for women aged 30–55 years, with testing done every 5 years. In the early 1990s, following an increase of the upper age limit to 60 years, this invitation-based system achieved an overall coverage of 75%. It has been estimated that this programme reduced mortality from cervical cancer by 90%. Broadly similar programmes were introduced at approximately the same time in Sweden, Iceland and parts of Norway and Denmark. In these countries, however, the reduction of mortality from cervical cancer has been less evident. This is probably due to a narrower age range covered in Sweden (where age-standardised death rates for the age group 0–64 years fell by 78% in the period 1970–95, compared to 87% in Finland), and to a lower overall coverage in Norway and Denmark (where rates fell by 61% and 71%, respectively). Hristova and Hakama produced projections of the effectiveness, cost and cost-effectiveness of adopting the Finnish screening model in all Nordic countries. They concluded that in the period 1998–2002 the programmes would have an incremental cost-effectiveness varying between US$1000 and 2600 per QALY gained for Finland and Denmark, respectively. Another study conducted in Denmark showed that the national guidelines for cervical screening issued by the National Board of Health in 1986 advocated a suboptimal strategy, with an incremental cost-effectiveness of approximately DKK 113,500 (assuming 80% coverage) compared to the next less effective alternative strategy. The authors examined a wide range of screening options and concluded that the lowest cost-effectiveness ratio could be obtained by screening women aged 30–59 every 5 years (the incremental cost-effectiveness versus targeting women aged 30–50 years was DKK 20,500, the average cost-effectiveness was DKK 21,000).

In the late 1970s an organised screening programme was progressively introduced in The Netherlands, initially aimed at women aged 35–55 years, to be screened at 3-year intervals, and subsequently extended to cover the age range 30–60 years, but with screening at 5-year intervals. In recent years the programme has achieved a national coverage of 75% up to the age of 50 years. After the introduction of the national screening programme in The Netherlands a reduction in mortality of 58% was achieved, although it must be said that rates fell by 42% between 1970 and 1980, when screening was very limited. Using a mathematical model it has been estimated that the current screening strategy may prevent 4563 deaths in the period 1993–2020 and save 118,824 life-years at a cost-effectiveness ratio (discounted at a 5% rate) of DFL 27,602 per life-year saved (equivalent to £9858 in 1993), compared with a hypothetical situation in which screening were not available.

Outside Europe, a national screening programme was implemented in Japan in 1961, although 30 years later coverage was still as low as 15.4%. A simulation based on a hypothetical cohort of 200,000 women followed up for 40 years starting from the age of 30 years produced a cost-effectiveness estimate of US$40,604 per life-year saved (equivalent to £26,828 in 1990).

A very detailed study based on a simulation model was conducted by Eddy, mainly based on data extracted from the literature. The author estimated the reduction in the risk of developing cervical cancer due to screening at different intervals, and the associated benefits in terms of life-years gained and financial costs for the American population. A thorough analysis of alternative strategies involving different screening intervals and different age...
ranges led to the conclusion that screening may reduce mortality from cervical cancer by approximately 90%, adding 96 days to the average life-expectancy of a 20-year-old woman, at a cost-effectiveness ratio in the region of US$10,000 per year of life-expectancy gained.\textsuperscript{134}

Evidence of the cost-effectiveness of cervical cancer screening in the UK is limited.\textsuperscript{155–159} This is discussed later in this chapter in the light of the results of our analysis.

Cost-effectiveness analysis of the national programme in England

A mathematical model was developed for the purpose of assessing the cost-effectiveness of the screening strategy currently adopted in England compared with a hypothetical situation with no organised programme, and for assessing the cost-effectiveness of a possible further increase in coverage. The model makes use of secondary data selected from different sources (as illustrated in appendix 6). It presents an innovative structure compared to previous models for the analysis of the cost-effectiveness of cervical cancer screening,\textsuperscript{134,151,153,154} being based on a cross-sectional rather than longitudinal design. This methodological choice reduces significantly the complexity of the model and evaluates the cost-effectiveness of the programme for a population (rather than an individual or a cohort of identical individuals). The model is based on the assumption that the programme has been operating for a number of years, and therefore start-up effects are neglected.

The model evaluates the outcomes of alternative cervical screening programmes for women aged 20–64 years on the incidence of invasive cancer of the cervix and mortality in England, taking into account the spill-over effects in the age range 65–69 years.

Several sensitivity analyses were performed to assess the robustness of cost-effectiveness estimates to alternative assumptions about key variables.

The following screening strategies were examined and compared:

- **Option 1 – Hypothetical do-nothing.** In this scenario no screening would take place. The entire female population is assumed to have a risk of developing invasive cervical cancer equal to that of women not undergoing screening, calculated on the basis of 1997 age-specific incidence and coverage rates and relative risks by time elapsed from the last smear. Of course, this is a purely hypothetical situation and was included only for the purpose of assessing the overall impact of screening.

- **Option 2 – Strategy before target payments.** Before 1990, GPs received a flat fee per cervical smear. From the latest figures available before the introduction of target payments, a coverage of approximately 59% for screening at 5-year intervals was achieved for women aged 20–64 years in England.\textsuperscript{162} This option was taken into account mainly for the purpose of estimating the cost-effectiveness impact of the target payment strategy.

- **Option 3 – Current strategy with target payments.** This option reflects the adoption of the strategy currently in place, with a majority of districts inviting women for screening at 3-year intervals and a minority at 5-year intervals. The current share of the eligible population covered by screening in the last 3 years is 67%, whereas 16% of women have had a smear between 3 and 5 years ago and 17% remain unscreened.\textsuperscript{146}

- **Option 4 – 95% coverage.** This option is hypothetical and involves expanding coverage to 95% while maintaining the current proportions of women screened 3-yearly and 5-yearly.

The model produced estimates of the costs and effectiveness of each option based on data gathered from several sources of information. These are illustrated in detail in appendix 6.

Results of the cost-effectiveness analysis

The model indicates that dramatic improvements in mortality from cervical cancer have been achieved in England since 1990, in correspondence with the introduction of the current screening strategy and compared with a hypothetical situation in which screening were not available (775 cancer deaths averted and 12,076 life-years saved per year), or to maintaining the strategy in use until 1990 (272 deaths averted and 4158 life-years saved per year). Such improvements have been achieved at a cost-effectiveness ratio of £11,191 per life-year saved (£/year 2000), compared with the hypothesis of maintaining the strategy in use until 1990. Table 6 shows the results of this first comparison in terms of costs, outcomes and incremental cost-effectiveness. The current strategy entails a significant commitment of resources. There is wide agreement that cervical cancer screening is not a form of screening that may ‘pay for itself’ just by reducing treatment costs.\textsuperscript{134,154} Even the previous strategy, which involved comparatively modest payments to GPs,
had a higher cost than the (theoretical) cost of treating all cases at a symptomatic stage. However, as previously mentioned, the cost per life-year gained through the current strategy compares favourably with those of other interventions commonly used in European healthcare systems.

The results provided by the model seem to indicate that a further reduction in the incidence of invasive cervical cancer and related deaths may still be achieved at an acceptable cost-effectiveness ratio. It appears that increasing coverage while maintaining the current proportion of women screened at 3- and 5-year intervals would save an estimated 101 additional cancer deaths at a cost-effectiveness ratio that varies between £21,000 and 33,000 per death averted (versus the current strategy), depending on the cost of reaching additional women (varying in the range £10–25). This strategy, however, would involve an increase in NHS expenditure for the programme of up to £51 million, which may not be acceptable due to tight budget constraints. The robustness of these results has been confirmed by an extensive sensitivity analysis, the details of which are reported in appendix 6.

The results of our analysis are consistent with those of previous studies conducted in the UK. In a series of articles published in Cytopathology, Waugh and co-workers\textsuperscript{155–157} reported the findings of a thorough economic analysis of the screening programme in Tayside (Scotland). Their conclusion is that both switching to screening at 3-year intervals and extending coverage would give additional benefits at a favourable cost-effectiveness ratio (our figures differ due to the discounting of life-expectancy gains and the inclusion of an additional cost for reaching uncovered women). Our estimate of the cost-effectiveness of the current programme is significantly more favourable than that reported by Charny and co-workers\textsuperscript{158} before the start of the national organised programme, providing further evidence of the higher efficiency of the current strategy compared to the old system. Finally, a simulation-based study reported in 1986\textsuperscript{159} reported estimates of the outcomes and cost-effectiveness of a range of alternative screening options that are difficult to compare with those produced by our model because of differences in the strategies explored.

What does appear from a direct comparison is that screening women in the younger age groups is not as inefficient as Parkin and Moss\textsuperscript{159} suggested more than 10 years ago, mainly due to increases in cancer incidence in such groups.

### Distributional effects of cervical cancer screening in England

Overall, the initial evaluations of the target payment system have been positive, a position confirmed by the National Audit Office in their latest review of the screening programme:

The target payment system is providing a good incentive to achieve and maintain the overall coverage of the cervical screening programme. Nearly 70% of general practices reported that they found the target payment scheme instrumental in encouraging them to achieve and maintain high screening coverage.\textsuperscript{135}

However, the fact that 13% of health authorities in England still do not reach the targets means that a considerable proportion of women remain underscreened. The National Audit Office confirmed the existence of significant gaps in coverage:

Research indicates that some groups of women including some ethnic minorities, unskilled manual workers and those in poor economic circumstances are underscreened.\textsuperscript{155}

These findings have been confirmed by other empirical research. Studies carried out since the introduction of target payments have highlighted broadly the same profiles of non-participants.
Unscreened women generally tend to be from social classes 4 and 5 rather than 1 and 2, unemployed and living in deprived areas and overcrowded accommodation.

Both Brown and co-workers and Austoker, while acknowledging that some progress has been made, have underlined that more concerted effort is needed to increase coverage among women from lower social classes (especially amongst those living in inner city areas) if the gap between classes in cervical cancer morbidity is to be narrowed.

Research has indicated that for both childhood immunisation and cervical cytology there is still a great deal of variation in target attainment between practices. More specifically, practices located in affluent areas are significantly more likely to reach the high target than are those located in deprived areas.

It is extremely difficult to assess the distributional consequences of the current cervical screening programme and explore whether variations in coverage may have exacerbated inequalities in morbidity and mortality from cervical cancer between women belonging to different social groups. In particular, difficulties arise because recent accurate measures of inequalities in incidence of cervical cancer are not available and effects on mortality become apparent only after a long time. However, even if information on these effects were available an assessment of the causal influence of an uneven screening coverage on inequalities would be compounded by several concurrent determinants.

As part of this case study we made an attempt to estimate indirectly the distributional impact of the current screening strategy by measuring the shortfall in the number of cervical cancer cases detected in recent years compared with model projections based on incidence rates prior to the introduction of target payments and assuming that increases in coverage had been achieved uniformly in all social groups. The simulation model is the same as that used for assessing the cost-effectiveness of the screening programme.

The starting point of the simulation is the year 1990, the earliest time for which complete data about cervical cancer incidence and screening coverage by age group are available. In 1990, 2744 cases of invasive cervical cancer were diagnosed in women aged 20–64 years. On the basis of age-specific incidence rates and relative risk rates related to the time elapsed from the last screening test (as illustrated in appendix 6), the risk of developing cervical cancer was determined by age and screening status. These values were used to formulate projections of cervical cancer cases that would have been diagnosed in 1995 and 1997 if the increase in screening coverage had been achieved uniformly across population groups. Projections were based on the assumption that coverage rates have not increased among women who have had no sexual partners. Projected numbers were compared to actual cases of cervical cancer (based on actual incidence rates for 1995 and on extrapolations for 1997, as explained in appendix 6) and differences were attributed to uneven screening coverage. Due to the cross-sectional nature of the model, a direct link between screening coverage and incidence is assumed in any year. However, increases (or decreases) in screening coverage impact on incidence rates with some delay, and therefore the model provides reliable estimates only when coverage is relatively stable. This is the case for the mid-1990s, as coverage increased only marginally after 1992. Data for 1997, although based on extrapolated incidence rates, should provide a relatively unbiased estimate of the coverage–incidence relationship. In 1990, instead, screening coverage was on a sharply increasing trend that lasted from 1988 to 1992. For this reason, it could not be assumed that the 1990 incidence was a reflection of the 1990 coverage, and alternative scenarios had to be tested as described below.

In 1997, 1879 cases of invasive cervical cancer were diagnosed. Screening coverage in 1997 showed a 28% increase over that in 1990. If this additional 28% of the female population aged 20–64 years undergoing screening had been evenly distributed across social groups, the number of invasive cervical cancer cases diagnosed in 1997 would have been significantly lower than the actual number. In particular, the risks of developing invasive cancer for different population groups that would have to be applied in order to project the exact number of cases diagnosed in 1997 are those that assume that the 1990 incidence of cancer is a reflection of a stable 45% coverage of the female population aged 20–64 years. Although in 1990 the actual coverage was approximately 56%, this was significantly lower in 1989 (around 40%) and until 1988 (around 20%), and therefore it is unreasonable to assume that the 1990 incidence corresponds to a 45% stable coverage. In order to obtain a more precise estimate of the shortfall in the number of cancer cases avoided in relation to the increase in screening coverage between 1990 and 1997, we modelled the actual increase in coverage that occurred from 1988 to 1990 and obtained new
baseline risks for our projections. The model takes into account that a number of ‘latent’ cases of invasive cancer are diagnosed simply as an effect of an increase in coverage. As the increase in coverage was substantial between 1989 and 1990, this effect has a great influence on the results of the model projections, and therefore alternative assumptions on the size of this effect were tested.

The results of the new projections indicate that the number of cervical cancer cases avoided in 1997 due to the screening strategy introduced in 1990 is, at best, 85% of the number of cases that would have been avoided if coverage had increased uniformly across social groups, but the proportion is more likely to be closer to 60%. In theory, from these projections it would be possible to derive implicit weights reflecting the equity–efficiency trade-off of the cervical screening policy, but there are so many uncertainties and assumptions in the model that the meaning of such weights would be very limited. What is important is that the model seems to confirm that the performance of the screening programme has not been as good as it could have been, due to a disproportionate coverage of women at lower risk, particularly women from more affluent social groups. The uneven coverage may have increased inequalities in the incidence of cervical cancer between social classes, although this should be confirmed by empirical data that are not available at present.

**Discussion**

This study provides evidence that was not previously available about the cost-effectiveness of cervical cancer screening in England. Estimates are based on a cross-sectional model and rely on the best and most up-to-date information available on cervical cancer incidence and mortality, screening coverage at different time intervals and relative risk of developing invasive cancer by time elapsed since the last smear. However, extrapolations had to be made for missing data or data provided in different formats (e.g. different age intervals). Moreover, the validity of the estimates reported herein is subject to a large number of factors that may influence disease and survival trends.

The programme currently offered by the NHS entails a significant commitment of resources, particularly in terms of payments to GPs, which currently represent nearly half of the total cost of the screening programme. The benefits of such a programme are remarkable, and these have been achieved at a cost-effectiveness ratio that may be considered acceptable on the basis of common standards, although significantly higher than the £2500 per life-year saved calculated for breast cancer screening, which is ten times more common than cervical cancer in the UK. Even setting aside the issue of political viability, and under optimistic assumptions regarding the decrease in coverage that would follow, the financial savings that could hypothetically be achieved by discontinuing target payments would not justify the high toll in terms of additional cancer morbidity and mortality.

On the other hand, it appears that the development of the current screening strategy was mainly aimed at maximising the coverage obtainable with the resources invested. Although the resulting overall effectiveness and cost-effectiveness of the programme are good, the same, if not better, effects could have been obtained by covering fewer women, or by screening at longer time intervals, while providing a more even coverage across social groups (particularly women from lower social classes, who are penalised by the current strategy due to their lower uptake rates). This indicates a form of equity–efficiency trade-off, which has been addressed by policy-makers when shaping the national screening programme. The policy adopted reflects an implicit attitude to provide equal access to screening (where access is defined purely on the basis of supply side factors) and maximise coverage. If different definitions of access (to take into account demand side factors) and efficiency (maximising health gain rather than coverage) had been adopted, the results of the screening programme would have been very different, and even more so if a different equity principle (e.g. equal health gain across social groups) had been adopted.

A normative economic analysis incorporating equity weights to reflect, for instance, the higher morbidity and mortality of women from less affluent social groups, would indicate that the current strategy performs poorly on cost-effectiveness grounds. Alternatively, a positive approach aimed at illustrating the consequences of alternative policy options might have led to different choices and to a different public judgement about the current strategy.

Improvements in the performance of the screening programme could still be obtained by increasing coverage further. However, current levels of coverage in England are higher than in any other country, and therefore the cost of reaching additional women is likely to be high. Interventions would have to target women in the extreme age ranges (20–24 and 60–64 years), women from more
deprived areas and women from ethnic minorities. By targeting these groups, such interventions may produce desirable outcomes also in terms of a more equitable distribution of coverage. Further research is needed into the extent to which coverage can actually be increased in groups currently presenting lower uptake rates and on the costs involved.

The allocation of kidneys for renal transplantation in the UK

A person with end stage renal disease (ESRD) will normally die within 6 months unless they are provided with renal replacement therapy. Haycox and Jones\textsuperscript{169} reported that in December 1990 an estimated 19,000 people in the UK were receiving some form of renal replacement therapy. Of these, approximately 53% were living with a transplanted kidney. Transplantation is therefore an important part of renal replacement therapy in the UK. The number of people who could potentially benefit from renal replacement therapy has been rising as a result of two factors:

- a greater willingness to treat certain groups of people previously denied treatment, such as people with significant co-morbidities or old people
- demographic change leading to an increase in the stock of people requiring renal replacement therapy.

A recent study has simulated the demand for renal replacement therapy over the next 15 years.\textsuperscript{170} The results predicted a rise of between 50% and 100% in the number of people requiring renal replacement therapy. There were also predicted to be more elderly people with co-morbidity under treatment. Given the shortage of kidneys, meeting this rise in demand will require an expansion in dialysis. It may also increase the number of people on the waiting list for a kidney transplant.

Transplantation is seen as offering patients an improved quality of life in comparison with that offered by dialysis. It may also offer survival benefits. Cost-effectiveness studies have tended to show transplantation to be less costly. However, kidneys for transplantation are in short supply. This creates the necessity for setting priorities between those who might benefit from transplantation. Since July 1998 this has been done in the UK via the New Kidney Allocation Scheme developed by the Users’ Kidney Advisory Group HLA Task Force of the UK Transplant Support Service Authority (UKTSSA).\textsuperscript{171}

The aim of this case study is to evaluate the New Kidney Allocation Scheme in the light of evidence of the effectiveness and cost of transplantation, to reveal the distributional implications of the scheme. There were three objectives:

- to determine the extent to which the scheme is supported by evidence of cost-effectiveness
- to determine which concepts of equity might be consistent with the scheme
- to evaluate whether the allocation scheme results in a trade-off between equity and efficiency.

The New Kidney Allocation Scheme in the UK

Since 1 July 1998 a new national donor kidney allocation scheme has been in operation in the UK. The scheme is coordinated by the UKTSSA and involves the cooperation of all renal transplant centres in the UK.

When a cadaveric donor kidney becomes available for transplantation, a list of patients with compatible blood types is prepared from the national database of patients waiting for a kidney transplant. These potential recipients are then organised into ranked categories based on four criteria:

- matchability
- recipient age
- sensitisation (those in matchability tier 1 only)
- location.

Thus, to determine to whom a kidney should be offered, the scheme first considers matchability. Amongst those equally ranked according to matchability, the scheme next considers recipient age. If two or more patients are still ranked equally, and are in matchability tier 1, sensitisation becomes the deciding criterion followed by location. If ranked equally, but in lower matchability tiers, location determines priority. If, having gone through this process, a ‘winner’ has not been selected, further criteria apply. In the case of paediatric patients (age 0–17 years) waiting times are used. For adults (age 18 years and above) a point-scoring system is applied. This allocation scheme is described in detail below.

**Matchability**

Matchability refers to the human leucocyte antigen (HLA) complex of genes which determine the histocompatibility complex of each individual. This is important to how an individual’s immune system responds to a transplanted organ and thus to the
likely success of a transplant. HLA genes are divided into three types: HLA-A, HLA-B and HLA-DR. These are used in the allocation scheme in the following way.

Patients are divided into three tiers in order of priority:

- Tier 1 – patients with no mismatches in either HLA-A, HLA-B or HLA-DR.
- Tier 2 – patients with other favourable matches (defined as no more than one mismatch for HLA-A and/or HLA-B and no mismatches for HLA-DR).
- Tier 3 – patients with non-favourable matches.

A patient in tier 2 will only be offered the kidney if there are no patients in tier 1, or if, for some reason, the kidney is declined by patients in tier 1. Similarly, a patient in tier 3 will only be offered the kidney if there are no patients in tier 2, or if the kidney is declined by patients in tier 2. Note that in the case where there are no tier 1 patients either locally or nationally, a transplant centre is only required to make one of a pair of kidneys nationally available. The other kidney can be kept for use locally. If there are no tier 1 or tier 2 patients, both kidneys can be retained locally. Within tier 1, patients are subdivided into three further hierarchical priority categories: firstly by age, then by sensitisation, then by location. In tier 2, patients are subdivided by age and then location.

Recipient age
Patients are divided into two categories according to recipient age: 0–17 years (paediatric) and 18 years and older (adults). Patients defined as paediatric are given priority over patients defined as adult. Therefore, a patient who is 18 years or over can only receive a kidney if there are no compatible patients in the same matchability tier who are below 18 years of age.

Sensitisation
Within matchability tier 1, highly sensitised patients are given priority over non-highly sensitised patients. (Sensitised patients are in a state of hypersensitivity to an allergen. If they encounter this allergen, for example, in the form of a transplanted organ, they produce a strong immune response against it; in the case of a transplanted organ, this might result in rejection.)

Location
Within all matchability tiers, patients defined as ‘local’ (a patient at the transplant centre where the kidney has been recovered) are given priority over patients defined as ‘national’ (all other transplant centres).

Additional criteria
The above system is supplemented by additional criteria to discriminate between equally ranked patients.

Deciding ties between equally ranked paediatric patients
If two equally ranked patients are identified, the patient who has been waiting the longest is offered the kidney. If two equally ranked patients have the same waiting time, the patient nearest in travel time to the offering centre is prioritised.

Deciding ties between equally ranked adult patients
A point-scoring system is used to prioritise equally ranked adult patients. The patient with the highest point score is offered the kidney. Points are awarded on the basis of:

- recipient age (1–10 points; favours younger patients)
- donor–recipient age difference (1–10 points; avoids large age differences)
- waiting time (0.5–5 points; favours longer waiting)
- matchability (1–10 points; favours rarer HLA types)
- sensitisation (0.5–3.5 points; favours low sensitisation, avoids positive cross-matches)
- balance of exchange (1–10 points; favours higher centre balance).

Implications of the scheme for efficiency and equity

Matchability
Matchability would appear to be an efficiency criterion, since transplants with favourable matches are more likely to be successful. This indicates that efficiency is an important consideration within the new allocation scheme.

However, there may be implications for equity, stemming from the emphasis on matchability. There is a danger that the primacy given to matchability within the allocation scheme might discriminate against people with rare HLA types. As noted in Morris and co-workers, this may particularly affect patients from ethnic minorities. In the case of patients aged 18 years and above, the points scheme counterbalances this by awarding points to favour of rarer HLA types (1–10 points). This may be interpreted as a principle of fair opportunity to benefit.
Sensitisation
The scheme prioritises highly sensitised patients over non-highly sensitised patients. Is this based on a judgement of health gain or on a judgement of fairness? On the one hand, the highly sensitised person has a greater chance of rejecting the kidney. Therefore, more health can be gained by offering the kidney to the non-highly sensitised person. However, over time, this may not be true. The highly sensitised are only compatible with exactly matching donor kidneys; non-highly sensitised patients can be transplanted with a greater range of HLA types. Therefore, prioritising highly sensitised patients may maximise the gain available from a given flow of donor kidneys by allowing transplantation of less favourable matches in non-highly sensitised patients.

However, priority for the highly sensitised may also reflect a principle of fair opportunity to benefit (or fair access to treatment) from transplantation. Highly sensitised patients are suitable for a narrower range of kidneys and therefore have a lower probability of being compatible with any given donor kidney that becomes available. This may be seen as unfair. Giving priority to the highly sensitised might improve fairness in access between non-highly and highly sensitised patients.

Location
The time which the kidney has spent outside the body is inversely related to successful outcome. The use of location for patients equally ranked by matchability, age and sensitisation would seem to be a criterion based on expected transplant outcome.

Recipient age
There are three ways in which recipient age acts as a criterion for determining priority. First, priority is given to those aged 0–17 years. The allocation scheme therefore places greater value on the health of 0–17 year olds than on the health of those aged 18 years and over (given equal matchability).

Secondly, age appears as a criterion in the points scheme for determining priority between equally ranked people aged 18 years and over. Points are awarded on the basis shown in Table 7. Thus, someone aged between 18 and 27 years receives ten times the number of points of someone aged 64 years or over, and twice as many points as someone aged between 46 and 49 years. However, recipient age receives the same weight within the overall points scheme as does donor–recipient age difference, matchability and balance of exchange. The maximum possible overall points score is 48.5. Recipient age accounts for 10 of these.

There is a third possible route by which recipient age may influence the age distribution of kidneys. The scoring of points awarded for the donor–recipient age difference is designed to favour smaller differences in age between the donor and the recipient. Donor age has been found to be an important predictor of kidney transplant survival. Transplanted kidneys from young donors have significantly improved survival in comparison to kidneys from older donors, regardless of recipient age. The age-matching element of the point scheme increases the chances that a kidney from a young donor will be offered to a young recipient, and that a kidney from an older donor will be offered to an older recipient. This might reflect a belief that greater benefits are derived by transplanting younger people rather than older people with the kidneys that are likely to last longest. It might also reflect the view that it is fairer to give to younger people the kidneys most likely to result in successful transplant.

In synthesis, although there are a number of possible distributional implications of the scheme, including implications relating to rarer HLA types and the degree of sensitisation, in the rest of this case study we focus on the distributional implications of the treatment of recipient age. We have seen that the scheme prioritises younger over older people. In the next section, we assess whether this is supported by evidence that kidney transplantation is more effective in younger age groups.

Evidence of the effectiveness and cost of renal transplantation
Survival
A recent UK study looked at survival in different age groups following transplantation. A total of 6363 cadaveric renal transplants in people aged

<table>
<thead>
<tr>
<th>Age band (years)</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>18–27</td>
<td>10</td>
</tr>
<tr>
<td>28–32</td>
<td>9</td>
</tr>
<tr>
<td>33–36</td>
<td>8</td>
</tr>
<tr>
<td>37–41</td>
<td>7</td>
</tr>
<tr>
<td>42–45</td>
<td>6</td>
</tr>
<tr>
<td>46–49</td>
<td>5</td>
</tr>
<tr>
<td>50–53</td>
<td>4</td>
</tr>
<tr>
<td>54–57</td>
<td>3</td>
</tr>
<tr>
<td>58–63</td>
<td>2</td>
</tr>
<tr>
<td>64</td>
<td>1</td>
</tr>
</tbody>
</table>
18 years or older, carried out between 1986 and 1993, were analysed to determine factors affecting survival. The age of the recipient was found to be an important factor determining outcome (other factors included HLA matching, donor age, cause of donor’s death, cold ischaemic time and whether the recipient had diabetes). Transplant recipients were analysed in four age groups: 18–39, 40–49, 50–59 and 60 years or older. A quadratic model of risk of graft failure against recipient age showed risk falling slightly between age 18 years and the late-20s, before continuously rising from age 30 years upwards. However, this study made no comparison with survival on dialysis.

Vollmer and co-workers\textsuperscript{175} compared the survival experience of 1038 patients with ESRD treated with either transplantation or dialysis (or both). Controlling for the confounding effects of age and morbidity, they found that there was no significant difference in survival between dialysis and cadaveric transplantation.

More recent results have contradicted this finding. Schnuelle and co-workers\textsuperscript{174} in a European study of 309 patients, found that survival was improved by renal cadaveric transplantation. However, a large multicentre study from the USA\textsuperscript{175} has provided very strong evidence of the effectiveness of transplantation in terms of mortality reduction. This study used data from the US Renal Data System on 252,358 patients under the age of 70 years. Within the period of the study, 46,164 patients were placed on the waiting list for a first transplantation, and 23,275 of those on the waiting list received a first cadaveric transplant. The results showed that the annual death rate for all patients on dialysis was 2.6 times higher than that for patients on the waiting list, and the annual death rate for patients on the waiting list was 1.7 times higher than that for transplant recipients. Adjusting for age, sex, race and diabetes as the cause of renal disease, the relative risk of death across all subgroups for recipients of transplants compared with that for those on the waiting list was 0.66. Comparison of transplantation against the waiting list by age at time of initial placement on the waiting list showed relative risks of 0.33 for those aged 0–19 years, 0.24 for those aged 20–39 years, 0.33 for those aged 40–59 years and 0.39 for those aged 60–74 years at the time of being added to the list.

Schaubel and co-workers\textsuperscript{176} analysed data on 6400 patients aged 60 years and over for survival with transplant or on dialysis. The results indicated that elderly patients who received a transplant experienced significantly greater survival probability than did those who remained on dialysis. The 5-year survival rates were 81% and 51% for the transplant and dialysis groups, respectively. This result suggests significant extra survival benefits for elderly people from transplant over dialysis.

The evidence about survival in children suggests that transplantation is not as successful in the very young as in older children. Briscoe and co-workers\textsuperscript{177} cite 11 reports in which a poor graft outcome was reported for children aged less than 2 years. In addition, analysis from a large, multicentre study in the USA (the North American Paediatric Renal Transplant Cooperative Study, which had 57 participating centres and 754 children with 761 transplants) concluded that a recipient age of less than 2 years is an independent risk factor for graft outcome.\textsuperscript{178} Further analysis of the data from this study\textsuperscript{179} showed that graft survival at 1 year in children transplanted at age 2 years or younger was 46%, whereas for all children (aged 0–17 years) graft survival was 74%. This study also showed mortality after 2 years to be higher in children transplanted when 1 year old or younger and in all children (16% and 8%, respectively).

In another large study, Pirsch and co-workers\textsuperscript{180} studied graft survival in 526 patients aged 1–55 years. They reported that recipient age was a significant predictor of graft loss, with younger recipients being at greatest risk.

Some studies have found evidence that contradicts this pattern. Briscoe and co-workers\textsuperscript{177} compared graft and patient survival in children aged less than 2 years and children aged 2–18 years. Cadaveric graft survival 5 years post-transplant was 38% in those aged less than 2 years and 42% in those aged 2–18 years. However, patient survival in those transplanted when under 2 years of age was comparatively poor. After 5 years, survival was 70% in those aged under 2 years and 93% in those aged 2–18 years. Leichter and co-workers\textsuperscript{181} found very little difference in graft survival in children transplanted between the ages of 1 and 5 years and children aged 6–18 years (2-year survival 80% and 83%, respectively). However, both these studies were based on small samples.

The above studies of transplant survival in children did not involve a comparison with survival on dialysis. It is therefore difficult to draw conclusions about the relative survival benefit offered by transplantation. However, the 1-year mortality of chronic dialysis of children aged 0–1 year has been estimated as 22% and that in children aged...
Quality of life
The evidence strongly supports the view that transplantation offers a superior quality of life to that offered by dialysis. Churchill and co-workers\textsuperscript{184} applied the time trade-off technique to determine health state utilities of patients on dialysis and patients following renal transplantation. They found that transplantation was valued on a 0–1 scale at 0.84, hospital dialysis at 0.43, home dialysis at 0.49 and continuous ambulatory peritoneal dialysis (CAPD) at 0.56. Hart and Evans\textsuperscript{185} used the sickness impact profile to assess quality of life in 859 renal replacement patients. The scores indicated superior quality of life for transplant recipients over those on dialysis. Gokal\textsuperscript{186} reviewed the evidence of the quality of life in renal replacement therapy. The general picture that emerged was one of a greater quality of life enjoyed by transplant recipients. More recently, Waiser and co-workers\textsuperscript{187} compared quality of life in 612 patients undergoing transplantation (359) and dialysis (253). Life satisfaction scores were higher in transplanted than in dialysis patients. Dialysis patients were also more anxious and more depressed.

However, these studies failed to control for patient selection into the various modes of treatment. Studies that have attempted to control for differences in the mix of patients receiving each type of treatment have found a similar pattern of results. Evans and co-workers\textsuperscript{186} measured the quality of life in patients undergoing either dialysis or transplantation. After controlling for case mix, they found that transplant recipients had higher quality of life on three subjective measures. They also found that functioning and ability to work were higher for transplantation. Laupacis and co-workers\textsuperscript{189} conducted a study of survival and quality of life over a 2-year follow-up period, using patients as their own controls. A variety of quality of life measures was used, including the time trade-off method for assessing preferences over health states and the sickness impact profile. The mean time trade-off score was 0.57 before transplantation and 0.7 at 2 years after transplantation. The proportion employed also increased from 30% before transplantation to 45% 2 years later. Overall survival was 91% at 2 years. Laupacis and co-workers\textsuperscript{189} also compared quality of life after 1-year and 2-year survival separately in those below and those above 60 years of age. The time trade-off scores were very similar: a change from 0.57 to 0.75 for those aged less than 60 years and a change from 0.55 to 0.72 for those aged over 60 years.

Cost-effectiveness
There is very little high-quality information on the relative cost-effectiveness of various treatments for ESRD. This absence stems from the methodological problems of patient selection and imperfect substitution between treatment modalities.

Garner and Dardis\textsuperscript{190} calculated the cost per life-year gained for four ESRD treatments over 5, 10, 15 and 20 years. They concluded that home dialysis and transplantation of living-related donor kidneys were more cost-effective than both in-centre dialysis and transplantation of cadaveric donor kidneys. They calculated further cost-effectiveness in six age groups: 0–10, 11–20, 21–30, 31–40, 41–50 and 51–64 years. They did not present the full results of this in their paper, but they did comment that the lowest costs per life-year gained for those that receive a transplant were obtained for individuals aged 21–40 years.

Karlberg\textsuperscript{191} estimated the costs of dialysis and transplantation in Sweden. The mean annual costs per individual of the various modes were calculated to be: hospital dialysis, SEK 400,000; outpatient dialysis, SEK 225,000; CAPD, 200,000; and transplant, SEK 60,000 (combining living-related and cadaveric transplants). These costs can be interpreted as the cost per life-year saved, since patients will die in the absence of treatment. The net saving in Sweden per kidney transplanted was estimated to be SEK 3.8 million.

Croxson and Ashton\textsuperscript{192} compared CAPD, home haemodialysis, in-centre haemodialysis (in-centre H) and transplantation with the no-treatment option. The measured costs per year of life saved for each treatment modality after 5 years was: CAPD, NZ$26,390; home haemodialysis, NZ$28,175; in-centre haemodialysis, NZ$35,270; and transplantation, NZ$18,463 (combining living-related and cadaveric transplants). This study is more of a description of costs and outcomes than a comparison, since, as the authors made clear, patients were selected into the various treatment options: the treatments are not perfect substitutes for each other. (Neither the overall sample size nor the number of cadaveric or living-related transplants were reported.)
Laupacis and co-workers\textsuperscript{189} conducted a cost-effectiveness analysis of haemodialysis and transplantation (combining living-related and cadaveric transplantation). Effectiveness was measured as QALYs. They compared dialysis and transplantation as alternatives by measuring the quality of life and cost, both before and after transplantation. Eligible subjects were aged 19 years or older, had been on the waiting list for at least 3 months and had received a transplant during the study period. Full follow-up data were available for 76 patients. Results were analysed by two age groups: those 60 years of age and below, and those aged above 60 years. The results showed transplantation to be dominant in both age subgroups. Overall, in year 1 transplantation cost US$66,540 and pretransplant dialysis cost US$73,659; in year 2, transplantation cost US$27,474 and pretransplant dialysis cost US$70,869. Quality adjustments in year 1 were 0.65 for transplantation and 0.55 for dialysis, and in year two 0.62 and 0.51, respectively. Therefore, transplantation was dominant.

The quality of life results in the two age groups are reported above. Interestingly, the costs both before and after transplant for those aged below 60 years were greater than the costs for those aged above 60 years.

**Discussion**

**Evidence of effectiveness**

The evidence described above strongly supports the case that there are survival benefits associated with transplantation in comparison with dialysis. These survival benefits would appear to hold across all age groups, including the very old. Several other age-related aspects of the effectiveness of transplantation are also apparent. There is good evidence that transplantation improves survival most effectively in young adults (Wolfe and co-workers\textsuperscript{175} found that the relative risk of death was lowest in the 20–39 year old age group). In addition, the estimates of expected lifetime from Wolfe and co-workers,\textsuperscript{175} based on extrapolation of their results, suggests that the incremental gain in life-expectancy in those aged 0–19 years is similar to that in middle-aged people. (If proved to reflect the future pattern of mortality in these patients, this is an interesting finding. It should be noted, however, that this is an extrapolation of relative risks estimated at 18 months, and that the confidence interval around the estimate of relative risk in those aged 0–19 years, although statistically significant, was quite large, indicating an imprecise estimate. Interpretation of the relative risks should also take into account the possibility that part of these reductions in risk may be explained by the selection of healthier patients for transplantation.) Evidence to support the finding by Wolfe and co-workers that young adults have the greatest potential to gain from transplantation was found by Morris and co-workers.\textsuperscript{172} They found that graft survival peaked in people in their late-20s. (Children were not included in this study, but a quadratic model of graft survival and age showed mortality in people in their late-20s to be lower than at younger ages.)

It also seems that transplantation has a poorer outcome in very young children than in older children. Comparisons with dialysis which would show the relative benefits of transplantation in children of different ages are not currently very conclusive.

The quality of life evidence also favours transplantation, on both subjective and functional measures. Empirical studies have tended not to investigate quality of life in different age groups. The one study reviewed here which has (Laupacis and co-workers\textsuperscript{189}) found that there were similar improvements in those above and below 60 years of age.

**Evidence of cost-effectiveness**

In terms of cost-effectiveness, measured as the cost per year of life gained through treatment versus no treatment (comparison of average ratios), the pattern emerging from these studies is that transplantation has the lowest ratio and in-patient hospital care has the highest ratio, with other forms of dialysis lying in between. However, as noted in the presentation of the evidence, these studies do not control for patient selection. They are descriptions of the costs per year of treatment (survival) in each modality, given the selection of patients into each modality. Two of these studies contained some age-related analysis. Garner and Dardis\textsuperscript{190} found that, among those who received a transplant, the cost per year of life gained was lower in those aged 21–40 years than in younger or
older age groups. Laupacis and co-workers\textsuperscript{189} found that costs were slightly lower in those aged above 60 years in comparison with those aged below 60 years. The 2-year survival was 94\% and 75\% in those aged below and above 60 years, respectively, with similar quality-of-life scores as measured by the time trade-off.

**Equity, efficiency and the New Kidney Allocation Scheme**

The picture emerging from the review of the evidence of cost-effectiveness and effectiveness is one in which transplantation is dominant in all age groups. However, kidneys are in short supply, which implies that priorities have to be set. Application of the cost-effectiveness criterion would mean prioritising those patients expected to gain most per unit cost. The evidence we have reviewed suggests that the extra benefits of transplantation are greatest in those aged 20–39 years. This age group was estimated to gain an additional 17 years of life. Those aged 0–19, 40–59 and 60–74 years were estimated to gain an additional 13, 11 and 4 years of life, respectively.\textsuperscript{165} However, the cost savings available from transplanting younger patients are greater than the savings from transplanting older patients, since, in the absence of transplant, the younger patients would spend longer on dialysis. This suggests that transplantation in children dominates transplantation in people aged over 40 years (greater cost savings and similar additional survival). However, the situation is less clear in comparison with transplantation in those aged 20–39 years. Using survival as the outcome measure, the greater benefits available in transplanting 20- to 39-year olds are only obtainable at extra cost (reduced savings). Quality adjusting survival might change this picture. The longer life expectancy of those aged 0–19 years would, when quality adjusted, add more to the relative benefits of transplantation in this group.

In the light of this discussion, it seems there is no clear justification in terms of efficiency (defined as the maximisation of health gain) for the priority ranking in favour of those aged 0–17 years in the UK’s New Kidney Allocation Scheme. However, neither is this ranking necessarily in conflict with efficiency. Maybe the extra savings could be shown to outweigh the foregone benefits (opportunity costs). However, on current evidence it is not clear that transplanting children provides a better choice on cost-effectiveness grounds than does transplanting young adults. However, rationing in favour of those aged 0–17 years may be supported by an alternative principle, namely equity. The discussion of age weights (see pages 27 and 28) revealed the complexity of the role of age in fair rationing (for a discussion of equity in relation to age and a review of the empirical literature, see pages 23–26). While in many cases this empirical literature revealed a preference for younger individuals, this was not always true when choosing between young people. Some of the studies found a willingness to prioritise older children over younger children (see, e.g., Lewis and Charny\textsuperscript{100} and Busschbach and co-workers\textsuperscript{99}). This evidence would not seem to support the UKTSSA priority criteria. A more likely source of support is the fair-innings argument. Fair-innings arguments imply priority for younger people in general. As discussed in chapter 2 (see page 12), there are broadly two types of fair-innings argument: those that specify a threshold age and those that do not. If a threshold is specified, those above the threshold are given lower priority than those below it. However, the age of 17 years is below what would normally be considered a suitable threshold fair-innings. Thus a threshold type of fair-innings argument does not support the ranked priority given to those aged 0–17 years. What about fair-innings arguments without a threshold age? Williams\textsuperscript{69} suggests deriving weights which would reflect the extent to which an individual was below or above the fair-innings. Under such a system, those aged 0–17 years would receive a higher weighting, since they are farther from the fair-innings. However, the priority scheme gives equal weight to those aged 0–17 years, which is not consistent with this type of weighting.

Both the points awarded for recipient age and the age-matching element within the points scheme are broadly consistent with efficiency. Younger adults derive more health benefits from transplantation than older adults, and the savings related to avoided years of dialysis are greater when transplantation is offered to younger patients. In terms of equity, these elements of the points scheme are broadly in line with the pattern found in the empirical studies, in which there was a declining value of health from middle age onwards. However, in the point scheme the points awarded for recipient age begin to decline from age 28 years onwards. Cropper and co-workers\textsuperscript{116} found health at 30 years of age to be valued above health at age 20 years. Williams\textsuperscript{114} found that ‘when bringing up children’ was a time of life at which those surveyed thought it was especially important to be healthy. This casts doubt on whether the pattern supported empirically would begin declining from an age as young as 28 years (see the points scheme for recipient age on page 56).
We have discussed the age-related aspects of the New Kidney Allocation Scheme for the allocation of cadaveric donor kidneys in the UK. In general, the scheme incorporates a preference for younger people over older people. The evidence of the cost-effectiveness of transplantation in different age groups would seem broadly to support the role of age in the points scheme for determining priority between patients aged 18 years and over. Equity considerations also support this points scheme. However, the basis for ranking those aged 0–17 years above those aged 18 years and above is less clear, both in terms of efficiency and the views of equity expressed within existing empirical studies. This ranking is the most influential way in which age enters the allocation scheme. Further work is necessary to clarify whether those aged 0–17 years represent a better investment in terms of cost-effectiveness than do young adults. The basis for this ranking in terms of equity remains obscure.

**Screening for haemoglobinopathies in the UK**

Sickle cell disease is caused by an inherited abnormality of haemoglobin function caused by a point mutation in the $\beta$-globin gene. The sickle gene is found mainly in people of African, Mediterranean, Middle Eastern and Indian ancestry. Haemoglobin disorders follow a pattern of inheritance in which there is a 1:4 chance of an affected pregnancy if both parents are carriers of an abnormal gene. There are four main types of sickle cell disorders in the UK: sickle cell anaemia, sickle HbC disease, sickle cell $\beta$-thalassaemia, and sickle HbD disease. The most common is sickle cell anaemia. Sickle HbC disease tends to be less severe than the other forms.

Sickle cell disease causes significant morbidity and mortality. Zeuner and co-workers have reported that in the UK “the most common and important acute events include painful crisis, pneumococcal sepsis, splenic sequestration, acute chest syndrome, stroke and acute anaemia”. They report further that “Sickle cell disorders are increasingly becoming chronic diseases, causing, for example, renal failure and chronic lung disease”. Platt and co-workers calculated the median age at death for individuals with sickle cell disease. For children and adults with sickle cell anaemia, the median age at death was 42 years for males and 48 years for females; for those with sickle cell HbC disease the median age at death was 60 years for males and 68 years for females. Death rates from sickle cell disease are highest from the ages of 1–3 years. The predominant cause of infant death is infectious disease. Most of these deaths occur as a result of infection by streptococcal pneumoniae. In 1981, the Cooperative Study of Sickle Cell Disease (a multicentre collaboration in the USA) reported an annual incidence of pneumococcal septicaemia of 10 per 100 person-years of observation in children under 3 years of age, with a 30% case fatality rate.

Substantial evidence about the effectiveness and cost-effectiveness of sickle cell disease screening in the UK has been produced by two projects conducted within the NHS HTA programme. The aim of the present case study was to use such evidence as a basis for exploring the distributional implications of policy choices made in different areas in the UK. In practice, the cost-effectiveness models developed as part of the two previous projects were replicated and modified in order to assess the consistency of actual policies with alternative equity principles.

The main equity issue involved in this case is similar to that discussed earlier for cervical cancer screening, but the key equity dimension is ethnicity rather than socio-economic condition. As mentioned previously, the disease affects individuals from certain ethnic minorities to a much greater extent than others. The policy adopted in many areas of the UK is different from that adopted for cervical screening. Rather than trying to maximise coverage on the basis of a principle of equal access to testing, a selective screening policy has been preferred based on the identification and screening of births at higher risk. In other areas, where ethnic minorities represent a larger proportion of the resident population, a universal screening policy has been adopted, but this is probably more aimed at overcoming inaccuracies in the selection of high-risk births than at fulfilling an ideal of equal access to screening, as the gains in the low-risk population are negligible.

**Effectiveness of screening for sickle cell disease**

A screening programme requires both effective screening tests and effective treatment options. Screening newborn babies for sickle cell disease might be a worthwhile strategy if there are effective early interventions to combat infection. Two possible prophylactic treatment options are available to reduce the risk of pneumococcal infection: vaccination and prophylactic antibiotic treatment. However, both these interventions have
potential problems. It is known that young children have a poor antibody response, reducing the effectiveness of vaccination. Furthermore, not all pneumococcal serotypes are represented within vaccines. Vaccines are therefore of questionable efficacy. Penicillin prophylaxis, on the other hand, has possible problems relating to compliance, the possibility of infection with resistant pneumococci, and uncertainty about the longer term effect of prophylaxis on the naturally acquired immunity in early childhood.

However, a study carried out in the USA in the 1980s demonstrated the incremental effectiveness of prophylactic antibiotic treatment over vaccination in reducing the incidence of infection. Despite concerns about securing a sufficient level of compliance, prophylactic oral penicillin was shown to produce an 84% reduction in the incidence of infection. Most of those who acquired an infection had been vaccinated. Three deaths also occurred during the study: all three were in the control group and all three had been vaccinated.

An earlier study carried out in the UK had also shown the effectiveness of prophylactic antibiotic treatment in reducing infections in newborns with sickle cell disease in a comparison against vaccination (although without reaching statistical significance). This study found 11 cases of infection in the pneumococcal vaccine treated group; ten of these infections were by serotypes present in the vaccine.

These studies provide evidence that screening for sickle cell disease could be an effective health intervention. (However, it is worth noting that Gaston and co-workers point to two reasons for caution in interpreting their results: first, it is unknown how prophylactic antibiotic treatment might affect the development of future immunity; and, secondly, it is important to be aware of the possible adverse impact of prophylactic penicillin on the development of antibiotic-resistant organisms.) In the USA there have been policy recommendations that universal screening should be offered throughout the country. The NIH Consensus Development Conference on Newborn Screening for Sickle Cell Disease and Other Hemoglobinopathies concluded that “The benefits of screening are so compelling that universal screening should be provided”. In the UK, the Standing Medical Advisory Committee recommended that, pending further research, antenatal and neonatal screening should be universal in districts where over 15% of the population is from ethnic minorities. But whether or not screening should be provided, and what form the service should take, should depend on analysis of the implications for efficiency and equity of alternative policy options. To this end, we next consider the evidence relating to the cost-effectiveness of screening for sickle cell disease.

Cost-effectiveness of policy options for sickle cell disease screening

In this section we examine the results of five comprehensive studies of the cost-effectiveness of neonatal screening for haemoglobinopathies. Three of these studies are from the USA and two are from the UK. Other studies are reviewed in Zeuner and co-workers.

Tsevat and co-workers compared neonatal screening for sickle cell disease and the use of penicillin to treat infants found to have pneumococcal sepsis, with a no-screening strategy in which penicillin would be administered only once symptoms of sickle cell disease developed. Their model calculated the cost per life saved by the two strategies separately in three risk groups: black, non-black, low prevalence; and non-black, high prevalence. Data were obtained from the published literature. The costs per life saved in comparison with no screening were calculated to be US$3100 in the black population, US$1.4 million in the non-black, high prevalence population, and US$450 billion in the non-black, low prevalence population.

Sprinkle and co-workers calculated the cost per case of sickle cell disease found through universal and non-universal neonatal screening in each state in the USA. They compared this with the cost of detecting a case of phenylketonuria through universal screening. A chief concern of this study was to make recommendations on how the costs of screening in low-prevalence states could be reduced through sharing testing facilities with other states.

Gessner and co-workers developed a decision model to compare no-screening with universal or targeted neonatal screening. The model was first applied to the state of Alaska. It was found that, compared to no-screening, targeted screening would cost US$206,000 per death averted. The incremental cost of universal compared with targeted screening would be US$2,040,000 per extra death averted. However, the universal programme would prevent 50% more deaths than the targeted programme. The authors emphasised that the incremental costs of preventing deaths, of
both targeted screening in comparison with no-screening and of universal screening in comparison with targeted screening, will vary substantially between states. This is due to varying sickle cell disease prevalence, test costs, laboratory procedures and whether follow-up services are selective (of clinically significant traits only) or complete (including clinically insignificant traits).

There have been two recent UK studies.\(^{196,205}\) Davies and co-workers\(^{205}\) (also reported in Cronin and co-workers\(^{215}\)) compared the cost-effectiveness of universal and targeted neonatal screening for haemoglobinopathies. This study also compared two types of laboratory test (isoelectric focusing and high-power liquid chromatography). They discovered little difference in cost between the two types of test. The cost analysis revealed significant economies of scale-up to testing levels of 25,000 per year, and further economies available up to 40,000 to 50,000 tests per year. They calculated the incremental cost per case of sickle cell disease identified through a universal programme at varying disease rates and numbers of births.

Zeuner and co-workers\(^{196}\) developed a model to compare the cost-effectiveness of universal testing and selective testing based on maternal ethnic status. The model was applied to ethnic-composition data for district health authorities in the UK. They estimated that universal neonatal screening would cost approximately £22,000 per 10,000 antenatal population. Selective neonatal screening costs would range from £200 to £11,500 per 10,000 antenatal population. They concluded that screening neonates of North European women is not cost-effective, even under extreme assumptions about sickle trait prevalence and the rate of inter-ethnic union. The case for universal neonatal screening is that it will result in higher coverage of neonates born to ethnic-minority women.

**Distributional effects of sickle cell disease screening in the UK**

The cost-effectiveness analysis models recently developed in the UK\(^{196,205}\) indicate that a policy of selective neonatal screening is dominant over a policy of no screening, whereas a policy of universal screening is more effective and more expensive than a selective policy (although it may still be dominant over no screening), at a cost-effectiveness ratio that varies significantly with the proportion of high-risk births. However, the choice between universal and selective screening may have important distributional implications due to the strong association between ethnicity and risk.

In order to explore the distributional implications of alternative policies, we developed a simplified cost-effectiveness model, based on elements drawn from the two previously cited UK studies. The model considers only neonatal screening, because the assessment of antenatal (or combined antenatal and neonatal) screening is difficult due to the important ethical issues involved and because, as mentioned by Zeuner and co-workers\(^{196}\) and confirmed by data received from selected health authorities (see below), in many cases there is no transfer of information from antenatal to neonatal screening, which in practice makes the latter an independent programme. Screening for \(\beta\)-thalassaemia was not considered.

The model was used to determine the effectiveness and cost-effectiveness of selective and universal screening policies in different ethnic groups and in relation to the proportion of high-risk births. These were determined in terms of life-years gained and cost per life-year gained, respectively. Cost, risk and survival values were mostly derived from the study by Zeuner and co-workers,\(^{196}\) although neither this nor the other recent UK-based study reported cost-effectiveness ratios in terms of the cost per life-year gained. For illustrative purposes, we obtained data on the number of births, risk and ethnic mix from five areas in the UK (Birmingham, Cardiff, East London and City, Gloucester, Middlesex) in order to set our model estimates in specific contexts. Data from Zeuner and co-workers\(^{196}\) were used to fill gaps in the information provided by the health authorities.

In the first instance, we determined the incremental cost-effectiveness ratio of switching from a selective to a universal screening policy when suggested by the UK Standing Medical Advisory Committee\(^{208}\) (i.e. when the proportion of high-risk ethnic minority population is higher than 15%). In our model this percentage was referred to the number of births from mothers belonging to high-risk ethnic minority groups. The cost-effectiveness ratio is clearly influenced by the mix of ethnic groups present in a specific area as, for instance, the Asian population has a relatively lower risk compared to the black population. Using the baseline assumptions of the model developed by Zeuner and co-workers,\(^{196}\) the cost per life-year gained of universal screening versus selective screening with (hypothetical) 15% ethnic minority varies in our estimates from £430,000 to over £1 million, the former being referred to the ethnic mix of East London and City, where a relatively large proportion of the population is of black
African origin, and the latter to Birmingham, where Asian groups are dominant.

At a second stage, we looked at the real proportion of ethnic minority births in the five areas and determined the effectiveness of antenatal and neonatal screening in different ethnic groups and the cost-effectiveness of the policies adopted in the specific areas. In all cases, if selective screening targets babies born to all mothers of non-northern-European origin, as indicated by Zeuner and co-workers, the benefits of a policy of universal screening would still be almost exclusively for the group targeted by the selective programme. These benefits would be mainly related to the inaccuracy of selection procedures and to the problems caused by interracial marriage, which would be overcome by a universal programme. The white northern European group would receive very limited benefits from a universal policy, due to a very low incidence of the disease in this group.

In all areas selective screening was dominant over no screening. In two areas (East London and City and Middlesex) universal screening was also dominant over no screening. In low-incidence areas (Cardiff and Gloucester, both below the 15% threshold), where a selective screening policy has been adopted, universal screening would have a cost per life-year gained of almost £6 million and over £5 million pounds, respectively. In the other three areas, where high-risk ethnic minorities represent a larger proportion of the population (and of births), and where universal screening has been adopted, the cost per life-year gained by this policy versus selective screening varies from slightly over £40,000 for East London and City (59% ethnic minority, predominantly black), to almost £90,000 in Middlesex (53% ethnic minority, predominantly Asian), to almost £430,000 in Birmingham (31% ethnic minority).

Discussion

For illustrative purposes, we have produced estimates of the effectiveness and cost-effectiveness of alternative policy options for sickle cell disease screening with reference to five areas of the UK. These are summarised in the previous section. It must be emphasised that the results reported are to be interpreted with caution, because they are based on a simplified model that considers neonatal screening in isolation and because the data provided by health authorities was incomplete and had to be integrated with information from different sources. However, it must also be said that considering antenatal screening in association with the neonatal programme would most likely make our conclusions more extreme, as the incremental cost-effectiveness ratio of universal versus selective screening would appear even less favourable. The most important aspect of the results presented in the previous section is the order of magnitude of the cost-effectiveness ratios and the implications of these in terms of the equity–efficiency trade-off.

A policy of selective screening targeted at all births to women of non-northern-European origin is always dominant over no screening, but is to some extent discriminatory. What makes this policy appear acceptable in certain circumstances is that:

- those deliberately excluded from coverage belong to a social group that is generally healthier and wealthier than average
- the risk of developing the disease is very low in the population not covered by selective screening.

A switch to a universal screening policy would appear to dissipate any concerns about possible discrimination. In most cases, this entails bearing an extra cost over the no-screening option (universal screening is always more expensive than selective screening) at a cost-effectiveness ratio that varies significantly with the proportion of births at high risk. The high-risk group is still the main beneficiary of a universal screening policy, as the gains due to the elimination of the inaccuracies of the selection procedures are far greater than the gains in the low-risk population. The policy choices made in the UK may offer an indication of the value attached to the incremental benefits of universal screening, both in terms of additional health gain and in terms of more equal coverage (i.e. access, defined as in the cervical cancer screening case from a supply-side perspective), whenever financially viable.

Both the indication provided by the UK Standing Medical Advisory Committee (15% high-risk population threshold) and the policy choices made in three of the areas examined, where universal screening was adopted (Birmingham, Middlesex, East London and City), entail very high cost-effectiveness ratios versus the selective screening option, as compared with other interventions commonly provided by the UK NHS. These ratios seem to indicate that the policy-makers involved have a strong aversion to the exclusion of part of the population from the benefits of sickle cell disease screening (i.e. they tend to prefer a policy of equal access, in supply-side terms). For purely illustrative
purposes, if the policy choices made had to be justified on cost-effectiveness grounds, assuming that a ratio of £20,000 per life-year gained is closer to those of other interventions provided by the NHS, we would have to attach a weight over 20 times greater to health gains offered by universal screening to the group not covered by selective screening. As the incremental health gains of universal screening for the white northern-European group are still comparatively small, this does not appear to be primarily an ethnic issue. Providing equal access to screening, or equal access for equal need, seems to be the main concern of policy-makers in this case (as it seemed to be in the cervical cancer screening case study).

There is, of course, an alternative explanation to the policy choices made about universal and selective screening. Simply, the economic logic may have been applied incorrectly (or not at all) to the analysis of the consequences of such policies. If the cost-effectiveness of universal screening were determined in comparison with the no-screening option, this would appear much more favourable than the ratios presented in the previous section indicate. In particular, universal screening would be dominant in East London and City and Middlesex, whereas it would have a cost-effectiveness ratio of the order of £16,000 per life-year gained in Birmingham. The 15% high-risk population threshold would have a cost-effectiveness ratio no higher £50,000 per life-year gained. This represents a classical flow in cost-effectiveness calculations, which our review of economic evaluations (chapter 3) showed to be very common. When mutually exclusive interventions are available, a cost-effectiveness comparison requires ranking these in terms of their effectiveness and determining incremental cost-effectiveness ratios accordingly. Different comparisons would be misleading. The correct comparison in this case is between universal and selective screening, as the latter dominates the no-screening option but is less effective than a universal policy.
Chapter 5
Conclusions and recommendations for further research

The results of each of the three parts of the project have been presented and discussed. The purpose of this final section is to draw some overall conclusions from the work conducted and to indicate priority areas for further research.

The state of the art of healthcare economic evaluation, as illustrated by a systematic review of studies published in five sample years (1987, 1992, 1995, 1996, 1997), indicates that decision-makers operating at different levels in the NHS would not find in published cost–benefit and cost-effectiveness analyses a suitable guide for addressing equity concerns in resource allocation. Our review was mainly focused on the distributional effects of prioritising between interventions competing for a given pool of resources. The review shows that existing economic evaluations have not taken such effects explicitly into account, and they do not systematically provide the information that decision-makers would require to formulate for themselves a judgement on the desirability of alternative policy options. The review also indicates that other key distributional effects appear to be neglected by existing economic evaluations, namely those involved in switching between the (mutually exclusive) interventions compared in an evaluation and those involved in providing an intervention selectively to subsets of the population that may potentially benefit from that intervention.

Implementing the prescriptions of existing economic evaluations would penalise certain social groups that the NHS is supposed to cover and protect.

Many health economists have recently turned their attention to the development of methods for addressing equity concerns explicitly within economic evaluation. This is reflected in the vast amount of literature directly or indirectly relevant to this methodological issue, which we reviewed during the course of the project (as illustrated in chapter 2). However, this research area is relatively new. The methods so far suggested are still crude, and the evidence upon which normative statements about alternative distributions could possibly be based is at present very weak. The main question we have tried to address at this early stage through our methodological literature review is whether a normative approach based, for instance, on the adjustment of health outcomes by means of appropriate equity weights, would be desirable and consistent with the theoretical foundations of healthcare economic evaluation. We have identified a number of important problems that may limit the applicability of the normative approach, all of which require further theoretical and methodological research. These are summarised below.

Individual versus societal perspective in the assessment of distributional consequences

Concerns for equity are a result of two distinct but overlapping phenomena. First, individuals derive utility from other people's consumption of healthcare (externality of the interdependent-utility type) or simply from the availability of healthcare (externality of the option-value type). Second, the ethics upon which a society and its culture are based involve preferences over alternative distributions of health as well as wealth. There is a complex relationship between these two aspects. They overlap in part, but they may also lead to conflicting distributional judgements.
Ideally, an economic analysis aimed at addressing normatively equity concerns should take both wealth and health into account. However, disentangling the complex relationship between the two may be conceptually difficult. In an extra-welfarist perspective, non-health externalities may arguably be ignored, but it remains questionable whether distributional effects should matter at all in such a perspective.

**Consistency between interdependent utilities and outcome measures in cost-effectiveness analysis**

Interdependent utilities and other forms of externalities, such as option value, are typically non-health effects of healthcare. Whereas these can certainly be associated with utility measures adopted in cost–benefit analysis, the adjustment of non-preference-based health outcome measures (widely used in cost-effectiveness analysis) for non-health-related interdependent utilities in a normative framework may be questionable and may not lead to welfare improvements. The issue of whether QALYs are preference-based measures and reflect utilities is still controversial and requires further exploration.

**Empirical issues in the assessment of interdependent utilities and social-welfare functions**

There are significant methodological difficulties in the empirical assessment of interdependent utilities and of the values that determine social-welfare functions. Our review identified only one study suggesting an ‘admittedly crude’ method for eliciting interdependent utilities in healthcare. The literature on eliciting values underlying social-welfare functions is comparatively broader, but it still needs expanding, and the design and methodology of empirical studies need to be strengthened. The main methodological issues involved in eliciting the values upon which social-welfare functions may be based are the influence of framing effects and the multidimensional nature of equity judgements. Framing effects make an unbiased assessment of preferences over alternative distributions extremely difficult. It may be argued that an unbiased assessment should not be sought because actual distributional judgements are always influenced by contextual factors, but if this view is accepted, study results would not be generalisable in the definition of a social-welfare function. There are few and limited examples of multidimensional preference elicitation. Determining how different equity dimensions affect societal preferences over alternative distributions is difficult, but it is necessary if the resulting social-welfare function is to have normative strength. A further methodological issue is related to inferring a social-welfare function from preferences expressed by individuals, as many existing studies do. Setting aside the issues of sample size and representativeness, preferences expressed by individuals may arguably reflect interdependent utilities more than societal ethics. The ‘veil of ignorance’ approach might partly bridge the gap between individual and societal values, but to what extent this is true is still far from clear.

**Consistency between the maximands of economic analysis and policy-making**

Healthcare economic evaluations have been so far concerned with the maximisation of health gain (typically cost-effectiveness analysis) or with the maximisation of the utility that individuals derive from their own health gains (typically cost–benefit analysis). This implies that incorporating an equity dimension in economic evaluation in a normative framework would entail judgements over alternative distributions of health gain or health-related utility. However, the equity concerns that affect healthcare policy-makers’ decisions are not exclusively, and perhaps not primarily, related to distributions of health gain or health-related utility. Policy-makers may express equity concerns about access or about treatment. The case studies examined in this project, for instance, indicate that equality of access (crudely defined in supply-side terms) is a key consideration behind some health policy decisions. In a normative framework, the maximand of economic evaluation would have to be modified to reflect policy-makers’ concerns beyond health gain.

A prioritised research agenda would involve broadening the maximand of economic evaluation to include dimensions of concern to policy-makers (see above) as the most crucial issue to be addressed by methodological research, followed by a further investigation of the conceptual issues related to interdependent utilities and their relationship with societal ethics and the maximands used in economic evaluation. Finally, if it is
proven that interdependent utilities and social-welfare functions should be measured as part of the implementation of a normative approach to economic evaluation, empirical measurement issues should be explored in greater depth than they have been so far. Although largely conceptual and methodological, this research agenda has great relevance to decision-makers within the NHS, as it constitutes a prerequisite for enhancing the relevance of healthcare economic evaluations to the actual context in which services are planned and delivered.

However, the issues described above inevitably lead to uncertainty about the appropriateness of a normative approach to addressing equity concerns through economic evaluation, at least at the current state of development of economic evaluation methods and with the current knowledge of societal values about alternative distributions. The most obvious alternative to a normative approach is a positive framework within which information on the cost-effectiveness of health interventions would be provided alongside information on the distributional effects of such interventions, and the task of making an overall judgement about the desirability of alternative allocations would be entirely left to decision-makers. The information that should be provided by economic evaluations within such framework depends on the type of equity question that is to be addressed. The most important question is that concerned with the distributional effects of prioritising between interventions competing for a given pool of resources. In order to address this equity question, economic evaluations should in the first place be based on outcome measures that allow comparisons across interventions. They should describe in detail the characteristics of the populations that may benefit from alternative interventions, and should illustrate the effects of such interventions in different subgroups along any relevant equity dimensions. When interventions could be provided selectively to subgroups, provision to each subgroup should be considered as a separate intervention, and incremental cost-effectiveness ratios should be determined accordingly. This is simple and straightforward information that, nevertheless, is not normally provided by economic evaluations.

To complement the evidence gathered through systematic reviews of the methodological and empirical economic evaluation literature, we conducted three case studies of recent healthcare policies adopted in the UK. These provided evidence that supports some of the conclusions previously described. In particular:

- Two of the case studies, cervical cancer screening and sickle cell disease screening, confirm that access is a key concern for policy-makers. In both cases, policies of access maximisation and provision of equal opportunities to receive care have been pursued at a high price, in terms of a smaller health gain and, possibly, increasing health inequalities in the first case, and in terms of high cost-effectiveness ratios in the second case.
- The case study on donor organ allocation for renal transplantation confirms the difficulties involved in assessing the values underlying a social-welfare function. Even if simplified to include one single key dimension (age), the equity judgement formulated by a group of experts on behalf of society does not appear fully consistent with the results of surveys of individual preferences.
- The equity–efficiency trade-offs revealed by some of the policy choices examined are so different from what would be expected on the basis of the evidence provided by existing empirical studies that they lead to the suspicion that information on the cost-effectiveness and distributional effects of the interventions involved was either not available or inappropriately used.

**Recommendations for further research**

The three components of this project illustrate the difficulties involved in pursuing a normative route to addressing equity concerns in healthcare economic evaluation. The most typical example of such a normative solution is the use of explicit equity weights, proposed by some authors as a methodological way forward for economic evaluation. The conceptual and practical limitations of this solution have been discussed extensively. Nevertheless, the equity dimension of healthcare resource allocation problems can no longer be ignored by those who undertake evaluations, as these are beginning to play a key role in healthcare policy-making. On the basis of the evidence gathered and reviewed in this report, it would seem reasonable to suggest a two-stage development of healthcare economic evaluation, structured as follows.

**Stage 1: short-term perspective**

An immediate priority for healthcare economic evaluation is to put an end to the complete neglect of distributional effects that characterises existing studies. This aim could be pursued in the short
term by using the solutions that are already available and do not present conceptual and methodological problems. In particular, what we described as the ‘positive route’ to addressing equity concerns may be adopted in the short term, involving the following actions:

- Researchers undertaking economic evaluations should systematically collect and report information on the characteristics of the populations that may benefit from the health interventions appraised, information on the effects of the interventions in different subgroups and, where appropriate, information on the cost-effectiveness of the interventions in such subgroups.
- Forms of sensitivity analysis may be employed by authors of economic evaluations to determine what (hypothetical) equity weights would support alternative policy choices.
- Existing methodological guidelines for the conduct and the reporting of economic evaluations may be adapted to incorporate the above recommendations.

**Stage 2: medium- to long-term perspective**

In the longer term, a substantial investment should be made in further research aimed to determine whether a normative solution may become possible at some stage in the future, based on the prioritised research agenda that we have indicated in this chapter. In particular, there is great scope for further research in at least three areas:

- **Conceptual analysis** of the relationship between individual and societal preferences over alternative distributions, of the relationship between different maximands in healthcare policy, and of the nature of QALYs and other outcome measures used in cost-effectiveness analysis.
- **Methodological research** on ways of obtaining unbiased and multidimensional measures of societal values with regard to equity (possibly by means of approaches such as multi-attribute utility theory and conjoint analysis).
- **Empirical research** aimed at eliciting such values as a basis for determining appropriate social-welfare functions.

If substantial progress is made on these three fronts, answers will become available for many of the questions currently surrounding the use of normative solutions to incorporating an equity dimension in economic evaluation. However, it is important to remember that (positive) methodological solutions are already available. These should be adopted without delay in order to prevent further proliferation of studies that offer a partial and potentially misleading view of the effects of alternative resource allocations, based on a narrow concept of efficiency maximisation.
This study was commissioned by the NHS R&D HTA programme. The authors are indebted to the HTA referees for their perseverance in reading this report and the quality of their comments. The views expressed in this report are those of the authors, who are responsible for any errors.

Although not listed as co-authors of this report and not responsible for its conclusions, a number of researchers contributed to parts of the project. In particular, Marco Oradei contributed to the literature searches, the cervical cancer screening case study and the sickle cell disease screening case study; David McDaid contributed to the literature searches; Monique Mrazek contributed to the review of economic evaluations; and Elena de Hoghton and Susanne Doehr contributed to the cervical cancer screening case study. The authors are also grateful to Penny Babb, Diane Sheerman-Chase, Eugenia Cronin, Charles Normand and various correspondents at the UK Transplant Support Services Authority and Health Authorities in Birmingham, Cardiff, East London and City, Gloucester and Middlesex, for providing essential information for the conduction of the three case studies.
References

4. Petty W. Political arithmetic, or a discourse concerning the extent and value of lands, people, buildings, etc. London: Robert Clavel; 1699.
References


References


100. Lewis PA, Charny M. Which of two individuals do you treat when only their ages are different and you can’t treat both? J Med Ethics 1989;15:29–32.


171. United Kingdom Transplant Support Service Authority (UKTSSA), Special Health Authority of the NHS. New Kidney Allocation Scheme. 1999.


**Appendix 1**

**Electronic databases used**

**MEDLINE**

MEDLINE is the US National Library of Medicine’s premier bibliographic database covering the fields of medicine, nursing, dentistry, veterinary medicine, the healthcare system and the preclinical sciences. It contains bibliographic citations from over 3900 biomedical journals published in the USA and 70 other countries. MEDLINE contains over 9 million records dating back to 1966. It has worldwide coverage, but 88% of the citations in current MEDLINE are to English-language sources. MEDLINE contains the citations that appear in Index Medicus, as well as the citations of ‘special-list’ journals. Special-list journals include those indexed for the Index to Dental Literature and the International Nursing Index.

**EMBASE**

EMBASE is the electronic version of the Excerpta Medica database, produced by Elsevier Science Publishers BV and supplied by Ovid Technologies Incorporated, with support from the Joint Information Systems Committee of the Higher Education funding bodies. This covers all aspects of human medicine and related biomedical research, including comprehensive information on the following subjects: drugs and toxicology, clinical medicine, biotechnology and bioengineering, health affairs, psychiatry and forensic medicine. The database includes data from 3500 biomedical journals published in 110 countries, dating back to 1980. The database contains more than 5 million records and is updated monthly.

**HealthSTAR**

HealthSTAR is a bibliographic database providing access to published literature of Health Services Technology, Administration, and Research, and is produced cooperatively by the US National Library of Medicine and the American Hospital Association. HealthSTAR contains approximately 3 million citations from 1975 to the present. It includes relevant bibliographic records from MEDLINE, and additional, unique, records specially indexed for it from three sources:

- journal articles with emphasis on healthcare administration, selected and indexed on an ongoing basis by the American Hospital Association
- a collection of 9451 retrospective (1975–81) monographs, technical reports and theses from the National Health Planning Information Center
- journal articles, technical and government reports, meeting papers and abstracts, and books and book chapters on healthcare research, clinical practice guidelines and healthcare technology assessment, selected and indexed on an ongoing basis through the National Information Center on Health Services Research and Health Care Technology (NICHSR).

**PsycLIT**

PsycLIT(R) contains citations and summaries of the journal articles, book chapters and book literature in psychology, as well as psychological aspects of related disciplines, such as medicine, psychiatry, nursing, sociology, education, pharmacology, physiology, linguistics, anthropology, business and law. Journal coverage, spanning 1887 to the present, includes international material selected from more than 1300 periodicals written in over 25 languages. Current chapter and book coverage includes worldwide English-language material published from 1987 to the present. Approximately 50,000 references are added annually through quarterly updates.

**EconLit**

EconLit is an indexed bibliography of the world’s economic literature, produced by the American Economic Association. It includes coverage of over 400 major journals, as well as articles in collective volumes (essays, proceedings, etc.), books, book reviews, dissertations and working papers licensed from the Cambridge University Press Abstracts of
Working Papers in Economics. Coverage is from 1969 to the present.

PAIS

The PAIS International database is produced by the Public Affairs Information Service, Inc. It is a bibliographic index to the literature of public policy, social policy, and the social sciences in general. Journal articles (from over 1000 journals), books, government documents, statistical compilations, committee reports, directories, serials, reports of public, intergovernmental and private organisations and most other forms of printed literature from all over the world are indexed. Coverage is from 1972 onwards, and it is updated quarterly.

IBSS

The IBSS is provided by the Bath Information & Data Services and is supplied and owned by the British Library of Political and Economic Science. The database regularly covers 2600 selected journals and 6000 books per year. In addition, over 1000 journals per year are covered selectively. The IBSS contains citations going back to 1951, and covers four disciplines: anthropology, economics, political science and sociology. The database is updated weekly.
Appendix 2

Inclusion criteria for the review of economic evaluations

- Article concerned with health interventions directly aimed at improving the health of individuals or populations?
  - Yes
  - No
- Abstract available?
  - Yes
  - No
- Is there evidence of consideration of costs and outcomes of two or more alternatives? Obtain full article?
  - Yes
  - No
- Article describes original research?
  - Yes
  - No
- Costs explicitly evaluated?
  - Yes
  - No
- Costs available from existing sources?
  - Yes
  - No
- Evidence of effectiveness available from existing sources?
  - Yes
  - No
- Effectiveness measured quantitatively?
  - Yes
  - No
- Evidence of implicit 'do-nothing' alternative considered?
  - Yes
  - No
- Alternatives considered?
  - Yes
  - No
- Evidence of consideration of costs and outcomes of two or more alternatives?
  - Yes
  - No
- Article describes original research?
  - Yes
  - No
- Costs explicitly evaluated?
  - Yes
  - No
- Costs available from existing sources?
  - Yes
  - No
- Evidence of effectiveness available from existing sources?
  - Yes
  - No
- Effectiveness measured quantitatively?
  - Yes
  - No
- Evidence of implicit 'do-nothing' alternative considered?
  - Yes
  - No
- Alternatives considered?
  - Yes
  - No
- Cost-effectiveness or cost-benefit analysis?
  - Yes
  - No
- Cost–consequence analysis: dominance?
  - Yes
  - No
- Paper included
- Paper excluded
Appendix 3

Checklist for the review of economic evaluations
Year: ____________________________  Journal: ________________________________

Author: ____________________________  Country: ________________________________

Intervention: __________________________________________________________________________________

☐ Preventive  ☐ Curative  ☐ Diagnostic  ☐ Screening  ☐ Rehabilitation  ☐ Palliative

1. Type of analysis:

☐ Cost–benefit

☐ Cost-effectiveness

☐ Other: _______________________________________________________________________________________

2. Perspective adopted

☐ Third payer

☐ Provider

☐ Societal

☐ Patient

☐ Other: _______________________________________________________________________________________

3. Outcome measure(s)

☐ Quality-adjusted life-expectancy

☐ Life-expectancy

☐ Survival rate

☐ Cases detected

☐ Cases successfully treated

☐ Cases prevented

☐ Other: _______________________________________________________________________________________

4. Setting

☐ Policy decision

☐ Clinical decision

☐ Managerial decision

5. Reference population

☐ Selected group

☐ Define group: ________________________________

☐ Unselected group
6. Information on the characteristics of the population that may benefit from the intervention(s)

- Size
- Age profile
- Gender profile
- Socio-economic profile
- Ethnic group profile
- Other dimensions: __________________________________________________________________________

7. Information on the effects of the intervention(s) in different subpopulations

- By age group
- By gender
- By socio-economic group
- By ethnic group
- By other dimensions: ________________________________________________________________________

8. Methods for assessing distributional impacts

- Differential valuation of outcomes in different subgroups
- Comparison of cost-effectiveness ratios by subgroups
- Other methods
Appendix 4

Economic evaluations reviewed in this report

1987


1992

Allen UD, Read S, Gafni A. Zidovudine for chemotherapy after occupational exposure to HIV-infected


Seron V, Moatti JP, Muller F, Le Gales C, Boue A. [Cost-effectiveness analysis of prenatal screening of trisomy 21...


1995


1996


Danese MD, Powe NR, Sawin CT, Ladenson PW. Screening for mild thyroid failure at the periodic health examination: a decision and cost-effectiveness analysis. JAMA 1996;276:285–92.


Livartowski A, Boucher J, Detournay B, Reinert P. Cost-effectiveness evaluation of vaccination against...


**1997**


Lorenzoni R, Fattore G, Gensini G. [Evaluation of the cost-effectiveness of thrombolytic therapy in acute myocardial infarct using tissue plasminogen activator or...


## Appendix 5

### Selected characteristics of the economic evaluations reviewed

<table>
<thead>
<tr>
<th>First author</th>
<th>Country</th>
<th>Type of intervention</th>
<th>Study type</th>
<th>Outcome measure</th>
<th>Effects in different subpopulations</th>
<th>Comparison of CER by subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beck</td>
<td>USA</td>
<td>Management of symptomatic bifascicular block</td>
<td>CE</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labelle</td>
<td>Canada</td>
<td>Extracorporeal shock wave lithotripsy, percutaneous ultrasonic lithotripsy and standard surgical treatment</td>
<td>CE</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dittus</td>
<td>USA</td>
<td>Patient management alternatives after uncomplicated myocardial infarction</td>
<td>CE</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Erickson</td>
<td>USA</td>
<td>Endoscopy</td>
<td>CE</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oster</td>
<td>USA</td>
<td>Alternative approaches to prophylaxis</td>
<td>CE</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specker</td>
<td>USA</td>
<td>Pulmonary embolism and lung scanning</td>
<td>CE</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Altman</td>
<td>USA</td>
<td>Smoking cessation programmes</td>
<td>CE</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thoner</td>
<td>Norway</td>
<td>Intensive care</td>
<td>CE</td>
<td>3</td>
<td>Age (0–19, 20–49, 50–69, &gt;70 years), diagnosis group, organ system with pre-existing dysfunction</td>
<td></td>
</tr>
<tr>
<td>Barry</td>
<td>USA</td>
<td>Faecal occult blood screening for colorectal cancer</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brandeau</td>
<td>USA</td>
<td>Work-up of asymptomatic patients with a positive faecal occult blood test</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eddy</td>
<td>USA</td>
<td>Screening for colorectal cancer</td>
<td>CE</td>
<td>4</td>
<td>Starting screening age (40, 45, 50 years)</td>
<td>Age (&gt;40, &gt;50 years)</td>
</tr>
<tr>
<td>England</td>
<td>USA</td>
<td>Surgery or angioplasty for renal revascularisation</td>
<td>CE</td>
<td>4</td>
<td>Age (&lt;35, 36–50, 51–60 and &gt;60 years), disease type, gender</td>
<td>Age (&lt;50, &gt;50 years)</td>
</tr>
<tr>
<td>Garner</td>
<td>USA</td>
<td>End-stage renal disease treatments</td>
<td>CE</td>
<td>4</td>
<td>Age, gender</td>
<td>Gender</td>
</tr>
<tr>
<td>Laffel</td>
<td>USA</td>
<td>Coronary thrombolysis, reperfusion therapy</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Continued*
<table>
<thead>
<tr>
<th>First author</th>
<th>Country</th>
<th>Type of intervention</th>
<th>Study type</th>
<th>Outcome measure*</th>
<th>Effects in different subpopulations</th>
<th>Comparison of CER by subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oster</td>
<td>USA</td>
<td>Antihyperlipemic therapy</td>
<td>CE 4</td>
<td>Age (35–74 years)</td>
<td>Age (35–39, 40–44, 45–49, 50–54, 55–59, 60–64, 65–69, 70–74 years)</td>
<td></td>
</tr>
<tr>
<td>Devereaux</td>
<td>USA</td>
<td>Echocardiography and electrocardiography</td>
<td>CE 5</td>
<td>Disease severity (mild, moderate, severe)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Davis</td>
<td>USA</td>
<td>Control of measles outbreak</td>
<td>CE 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hay</td>
<td>USA</td>
<td>Two strategies for prevention of <em>Haemophilus influenzae</em> type B infection</td>
<td>CE 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manau</td>
<td>USA</td>
<td>Prevention of dental caries</td>
<td>CE 6</td>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patriarca</td>
<td>USA</td>
<td>Prevention of type A influenza infections</td>
<td>CE 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carter</td>
<td>USA</td>
<td>Breast cancer screening</td>
<td>CB (HC) 7</td>
<td>Pre- versus postmenopausal women, risk profile</td>
<td></td>
<td></td>
</tr>
<tr>
<td>De Neef</td>
<td>USA</td>
<td>Rapid tests for streptococcal pharyngitis</td>
<td>CB (HC) 7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scriver</td>
<td>Canada</td>
<td>Screening urine for neuroblastoma</td>
<td>CE 7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spital</td>
<td>USA</td>
<td>Renal transplant</td>
<td>CB (WTP) 7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1992

<table>
<thead>
<tr>
<th>First author</th>
<th>Country</th>
<th>Type of intervention</th>
<th>Study type</th>
<th>Outcome measure*</th>
<th>Effects in different subpopulations</th>
<th>Comparison of CER by subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capri</td>
<td>Italy</td>
<td>Single and multiple dose antibiotic treatment of lower uncomplicated urinary tract infections</td>
<td>CE 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cheung</td>
<td>Australia</td>
<td>Hormone replacement therapy in menopausal women</td>
<td>CE 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fendrick</td>
<td>USA</td>
<td>Screening and treatment of diabetic retinopathy</td>
<td>CE 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freedberg</td>
<td>USA</td>
<td>Optimal management strategies for HIV-infected patients presenting with cough or dyspnoea</td>
<td>CE 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hall</td>
<td>Australia</td>
<td>Mammography screening in Australia</td>
<td>CE 1</td>
<td>Age (45–54, 55–64, 65–69 years)</td>
<td>Age (45–54, 55–64, 65–69 years)</td>
<td></td>
</tr>
<tr>
<td>Hillner</td>
<td>USA</td>
<td>Chemotherapy in node-negative breast cancer</td>
<td>CE 1</td>
<td>Age (45, 60 years)</td>
<td>Age (45, 60 years)</td>
<td></td>
</tr>
<tr>
<td>Hillner</td>
<td>USA</td>
<td>Node-negative breast cancer and adjuvant chemotherapy</td>
<td>CE 1</td>
<td>Age (45, 60 years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hillner</td>
<td>USA</td>
<td>Autologous bone marrow transplantation in metastatic breast cancer</td>
<td>CE 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johannesson</td>
<td>Sweden</td>
<td>Hypertension treatment</td>
<td>CE 1</td>
<td>Age, gender</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>First author</th>
<th>Country</th>
<th>Type of intervention</th>
<th>Study type</th>
<th>Outcome measure[^1]</th>
<th>Effects in different subpopulations</th>
<th>Comparison of CER by subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Larsen</td>
<td>USA</td>
<td>Implantable cardioverter–defibrillator</td>
<td>CE 1</td>
<td>Age (45, 55, 65 years)</td>
<td>Age (45, 55, 65 years)</td>
<td></td>
</tr>
<tr>
<td>Levin</td>
<td>Sweden</td>
<td>Intravenous recombinant tissue-type plasminogen activator in suspected myocardial infarction</td>
<td>CE 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mandelblatt</td>
<td>USA</td>
<td>Breast cancer screening</td>
<td>CE 1</td>
<td>Age (≥65 years), ethnicity</td>
<td>Age (60–69, 70–74, 75–79, 80–84, &gt;85 years)</td>
<td></td>
</tr>
<tr>
<td>Parker</td>
<td>UK</td>
<td>Hip-fracture treatment</td>
<td>CE 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teboul</td>
<td>France</td>
<td>Gallstone-removal strategies</td>
<td>CE 1</td>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uther</td>
<td>Australia</td>
<td>Automatic implantable defibrillator in prevention of sudden cardiac death</td>
<td>CE 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Egberts</td>
<td>Netherlands</td>
<td>Treatments for respiratory distress syndrome</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hillner</td>
<td>USA</td>
<td>Management of suspected lower-extremity deep vein thrombosis</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buxton</td>
<td>UK</td>
<td>Antiemetic agent (ondansetron)</td>
<td>CE 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haan</td>
<td>Netherlands</td>
<td>Effects of in vitro fertilisation</td>
<td>CE 3</td>
<td>Age (≥35 years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>King</td>
<td>UK</td>
<td>Rate of ambulatory peritoneal dialysis (APD) and continuous APD as long-term outpatient treatments for renal failure</td>
<td>CE 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kroll</td>
<td>USA</td>
<td>Mandibular reconstruction</td>
<td>CE 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee</td>
<td>USA</td>
<td>Outpatient-focused programme for childhood diabetes</td>
<td>CE 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lim</td>
<td>Singapore</td>
<td>Simvastatin and Gemfibrozil in the treatment of hyperlipidaemia</td>
<td>CE 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roquebrune</td>
<td>France</td>
<td>Therapeutic strategies for coronary artery disease</td>
<td>CE 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scott</td>
<td>USA</td>
<td>Intraoperative cholangiography</td>
<td>CE 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weissfeld</td>
<td>USA</td>
<td>Precision of blood cholesterol measurement and high blood cholesterol case-finding and treatment</td>
<td>CE 3</td>
<td>Age (50–60 years)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[^1]: Age ranges and subpopulations vary depending on the study and intervention type.
<table>
<thead>
<tr>
<th>First author</th>
<th>Country</th>
<th>Type of intervention</th>
<th>Study type</th>
<th>Outcome measure†</th>
<th>Effects in different subpopulations</th>
<th>Comparison of CER by subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilson</td>
<td>USA</td>
<td>Work-site cholesterol screening and intervention programmes</td>
<td>CE 3</td>
<td>Gender, occupational status, age (20–29, 30–39, 40–49, ≥50 years)</td>
<td>Age (50–64, 65–69, ≥70 years)</td>
<td>Age (50–64, 65–69, ≥70 years)</td>
</tr>
<tr>
<td>Brown</td>
<td>USA</td>
<td>Breast cancer screening</td>
<td>CE 4</td>
<td>Age (50–64, 65–69, ≥70 years)</td>
<td>Age (50–64, 65–69, ≥70 years)</td>
<td></td>
</tr>
<tr>
<td>Byers</td>
<td>USA</td>
<td>Screening for colorectal cancer</td>
<td>CE 4</td>
<td>Clinical status, age (50–54, 55–59, 60–64, 65–69, ≥70 years)</td>
<td>Clinical status, age (50–54, 55–59, 60–64, 65–69, ≥70 years)</td>
<td></td>
</tr>
<tr>
<td>Daly</td>
<td>UK</td>
<td>Hormone replacement therapy in menopausal women</td>
<td>CE 4</td>
<td>Clinical status, age (50–54, 55–59, 60–64, 65–69, ≥70 years)</td>
<td>Clinical status, age (50–54, 55–59, 60–64, 65–69, ≥70 years)</td>
<td></td>
</tr>
<tr>
<td>Desch</td>
<td>USA</td>
<td>Optimal timing of autologous bone marrow transplantation in Hodgkin’s disease after a chemotherapy relapse</td>
<td>CE 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Durand-Zaleski</td>
<td>France</td>
<td>Diagnosis of errors of amino acid metabolism</td>
<td>CE 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fahs</td>
<td>USA</td>
<td>Cervical cancer screening for the elderly</td>
<td>CE 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glazebrook</td>
<td>Canada</td>
<td>Radiation therapy</td>
<td>CE 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glick</td>
<td>USA and UK</td>
<td>Cholesterol-lowering therapy</td>
<td>CE 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goel</td>
<td>Canada</td>
<td>Intravenous tissue plasminogen activator versus streptokinase for acute myocardial infarction</td>
<td>CE 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johannesson</td>
<td>Sweden</td>
<td>Diet and drug treatment in obese men with mild hypertension</td>
<td>CE CB (WTP) 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Krumholz</td>
<td>USA</td>
<td>Thrombolytic therapy with streptokinase in elderly patients with suspected acute myocardial infarction</td>
<td>CE 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loissance</td>
<td>France</td>
<td>Therapeutic strategies in candidates for cardiac transplantation</td>
<td>CE 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O’Brien</td>
<td>UK</td>
<td>Implantable cardioverter–defibrillator</td>
<td>CE 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ramsay</td>
<td>USA</td>
<td>Prophylactic AZT following needlestick injury in healthcare workers</td>
<td>CE 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Siegel</td>
<td>USA</td>
<td>Screening and early treatment of nephropathy in patients with insulin-dependent diabetes mellitus</td>
<td>CE 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sonnenberg</td>
<td>USA</td>
<td>Preoperative testing of faecal occult blood</td>
<td>CE 4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>First author</th>
<th>Country</th>
<th>Type of intervention</th>
<th>Study type</th>
<th>Outcome measure^1</th>
<th>Effects in different subpopulations</th>
<th>Comparison of CER by subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calvo</td>
<td>USA</td>
<td>Low-osmolarity contrast medium for intravenous urography</td>
<td>CE 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Davies</td>
<td>UK</td>
<td>Duplex ultrasonography and pulse-generated run-off in selecting claudicants for femoropopliteal angioplasty</td>
<td>CE 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dervaux</td>
<td>France</td>
<td>Screening for colorectal cancer</td>
<td>CE 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heijboer</td>
<td>Netherlands</td>
<td>D-dimer test in combination with non-invasive testing versus serial non-invasive testing alone for the diagnosis of deep vein thrombosis</td>
<td>CE 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kang</td>
<td>Singapore</td>
<td>Strategies for hepatocellular carcinoma screening in hepatitis B virus carriers</td>
<td>CE 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lairson</td>
<td>USA</td>
<td>Alternative methods for diabetic retinopathy screening</td>
<td>CE 5</td>
<td>Ethnicity (white, Hispanic, black), age (&lt;65 years), education level (less than high school, high school, more than high school), employment status (full/part-time), monthly earnings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Niv</td>
<td>Israel</td>
<td>Questionnaire and colorectal screening</td>
<td>CE 5</td>
<td>Age (≥40 years), gender, ethnicity (Israel, Europe–America, Asia–Africa)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ruwe</td>
<td>USA</td>
<td>Magnetic resonance imaging versus diagnostic arthroscopy</td>
<td>CE 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sculpher</td>
<td>UK</td>
<td>Screening for diabetic retinopathy</td>
<td>CE 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sellors</td>
<td>Canada</td>
<td>Selective versus universal screening for chlamydial infection in sexually active young women</td>
<td>CE 5</td>
<td>Age (&lt;21 years), education, employment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seror</td>
<td>France</td>
<td>Prenatal screening for Down’s syndrome using maternal serum markers</td>
<td>CE 5</td>
<td>Age (≥35, ≥38 years)</td>
<td>Age (≥35, ≥38 years)</td>
<td></td>
</tr>
<tr>
<td>Turner</td>
<td>USA</td>
<td>Hearing screening protocols</td>
<td>CE 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Varenhorst</td>
<td>Sweden</td>
<td>Repeated screening for carcinoma of the prostate by digital rectal examination</td>
<td>CE 5</td>
<td>Age (50–54, 55–59, 60–64, 65–69 years)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>First author</th>
<th>Country</th>
<th>Type of intervention</th>
<th>Study type</th>
<th>Outcome measure(^1)</th>
<th>Effects in different subpopulations</th>
<th>Comparison of CER by subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walker</td>
<td>UK</td>
<td>Filtering strategies in mass population screening for colorectal cancer</td>
<td>CE</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whynes</td>
<td>UK</td>
<td>Screening strategies for colorectal cancer</td>
<td>CE</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zavertnik</td>
<td>USA</td>
<td>Management of breast cancer</td>
<td>CE</td>
<td>5</td>
<td>Ethnicity</td>
<td></td>
</tr>
<tr>
<td>Allen</td>
<td>Canada</td>
<td>Zidovudine for chemoprophylaxis after occupational exposure to HIV-infected blood</td>
<td>CE</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antonanzas</td>
<td>Spain</td>
<td>Hepatitis B vaccination</td>
<td>CE</td>
<td>6</td>
<td>Risk group</td>
<td>Risk group</td>
</tr>
<tr>
<td>Barrett</td>
<td>Canada</td>
<td>Non-ionic, low-osmolarity radiopaque contrast media versus ionic, high-osmolarity agents during cardiac catheterisation</td>
<td>CE</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brandeau</td>
<td>USA</td>
<td>Screening women of child-bearing age for HIV</td>
<td>CE</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caro</td>
<td>Canada</td>
<td>Replacing high-osmolarity with low-osmolarity contrast media</td>
<td>CE</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combs</td>
<td>USA</td>
<td>Autologous blood donation in obstetrics</td>
<td>CE</td>
<td>6</td>
<td>Ethnic group (Hispanic, Asian, other), prior abortions</td>
<td></td>
</tr>
<tr>
<td>Gustafsson</td>
<td>Sweden</td>
<td>Primary stroke prevention in atrial fibrillation</td>
<td>CE</td>
<td>6</td>
<td>Age (50–59, 60–69, 70–79, 80–89 years)</td>
<td></td>
</tr>
<tr>
<td>Jonsson</td>
<td>Sweden</td>
<td>Misoprostol for gastric ulcer prophylaxis</td>
<td>CE</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Levy</td>
<td>France</td>
<td>Flu vaccination</td>
<td>CE</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lieu</td>
<td>USA</td>
<td>Outpatient strategies for management of febrile infants</td>
<td>CE</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steinberg</td>
<td>USA</td>
<td>High-osmolarity compared with low-osmolarity contrast media in patients undergoing cardiac angiography</td>
<td>CE</td>
<td>6</td>
<td>Gender, age (&gt;60 years)</td>
<td></td>
</tr>
<tr>
<td>Tormans</td>
<td>Netherlands</td>
<td>Hepatitis A prevention in travellers</td>
<td>CE</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demicheli</td>
<td>Italy</td>
<td>Mass vaccination against hepatitis B</td>
<td>CB (HC)</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nicholl</td>
<td>UK</td>
<td>Lithotripsy and open cholecystectomy as treatments for gall-bladder stones</td>
<td>CE</td>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

continued
<table>
<thead>
<tr>
<th>First author</th>
<th>Country</th>
<th>Type of intervention</th>
<th>Study type</th>
<th>Outcome measure(^1)</th>
<th>Effects in different subpopulations</th>
<th>Comparison of CER by subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1995</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boer</td>
<td>Netherlands</td>
<td>Changing age limit for breast cancer screening</td>
<td>CE</td>
<td>I</td>
<td>Age (50–100 years)</td>
<td>Age (50–100 years)</td>
</tr>
<tr>
<td>Bush</td>
<td>USA</td>
<td>Aminotransferase to prevent post-transfusion hepatitis</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colice</td>
<td>USA</td>
<td>Head computed tomography in patients with lung cancer without clinical evidence of metastases</td>
<td>CE</td>
<td>I</td>
<td>Age (&lt;50, ≥50 years), gender</td>
<td></td>
</tr>
<tr>
<td>Etchason</td>
<td>USA</td>
<td>Preoperative autologous blood to reduce risk of HIV transmission</td>
<td>CE</td>
<td>I</td>
<td>Operative group</td>
<td>Age</td>
</tr>
<tr>
<td>Ferrell</td>
<td>USA</td>
<td>Low-air-loss beds for treatment of pressure ulcers</td>
<td>CE</td>
<td>I</td>
<td>Age, gender, ethnicity (Caucasian, black, other)</td>
<td></td>
</tr>
<tr>
<td>Gabriel</td>
<td>USA</td>
<td>Misoprostol prophylaxis</td>
<td>CE</td>
<td>I</td>
<td>Age (&gt;60 years, all ages)</td>
<td>Age (&gt;60 years, all ages)</td>
</tr>
<tr>
<td>Gage</td>
<td>USA</td>
<td>Prophylaxis for stroke</td>
<td>CE</td>
<td>I</td>
<td>Risk group</td>
<td>Risk group</td>
</tr>
<tr>
<td>Glimelius</td>
<td>Sweden</td>
<td>Palliative chemotherapy in gastrointestinal cancer</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Henon</td>
<td>France</td>
<td>Autologous blood cell transplantation versus chemotherapy</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holtgrave</td>
<td>USA</td>
<td>Prevention of HIV infection</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jonsson</td>
<td>Sweden</td>
<td>Fracture prevention</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kalish</td>
<td>USA</td>
<td>Thrombolytic therapy in myocardial infarction</td>
<td>CE</td>
<td>I</td>
<td>Age (&lt;75, &gt;75 years)</td>
<td></td>
</tr>
<tr>
<td>Kellett</td>
<td>Ireland</td>
<td>Accelerated tissue plasminogen activator with streptokinase for treatment of suspected myocardial infarction</td>
<td>CE</td>
<td>I</td>
<td>Age (50–80 years)</td>
<td></td>
</tr>
<tr>
<td>Kent</td>
<td>USA</td>
<td>Perioperative imaging for carotid endarterectomy</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kerridge</td>
<td>Australia</td>
<td>Intensive care</td>
<td>CE</td>
<td>I</td>
<td>Disease severity</td>
<td>Disease severity</td>
</tr>
<tr>
<td>Kiberd</td>
<td>Canada</td>
<td>Screening to prevent renal failure</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>King</td>
<td>USA</td>
<td>Surgery for intracranial aneurysms</td>
<td>CE</td>
<td>I</td>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>Lawrence</td>
<td>USA</td>
<td>Magnetic resonance angiography in renal failure</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lea</td>
<td>Australia</td>
<td>Cochlear implant</td>
<td>CE</td>
<td>I</td>
<td>Children and adults (not defined)</td>
<td></td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>First author</th>
<th>Country</th>
<th>Type of intervention</th>
<th>Study type</th>
<th>Outcome measure</th>
<th>Effects in different subpopulations</th>
<th>Comparison of CER by subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maureen</td>
<td>Netherlands</td>
<td>Inhaled corticosteroids and bronchodilators in asthma and chronic obstructive pulmonary disease</td>
<td>CE 1</td>
<td></td>
<td>Age</td>
<td>Age</td>
</tr>
<tr>
<td>Norum</td>
<td>Norway</td>
<td>Chemotherapy in gastric cancer</td>
<td>CE 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plans Rubio</td>
<td>Spain</td>
<td>Antipneumococcal vaccination</td>
<td>CE 1</td>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ramsey</td>
<td>USA</td>
<td>Lung transplantation</td>
<td>CE 1</td>
<td></td>
<td>Age</td>
<td>Age</td>
</tr>
<tr>
<td>Shepard</td>
<td>Canada</td>
<td>Child vaccination</td>
<td>CE 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tsevat</td>
<td>USA</td>
<td>Captopril therapy after myocardial infarction</td>
<td>CE 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wimo</td>
<td>Sweden</td>
<td>Dementia care</td>
<td>CE 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wyatt</td>
<td>USA</td>
<td>Cochlear implant</td>
<td>CE 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yin</td>
<td>USA</td>
<td>Magnetic resonance angiography</td>
<td>CE 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feig</td>
<td>USA</td>
<td>Mammography screening</td>
<td>CE 2</td>
<td>Age (40–49, 50–59 years)</td>
<td>Age (40–49, 50–59 years)</td>
<td></td>
</tr>
<tr>
<td>Field</td>
<td>UK</td>
<td>Coronary heart disease treatments</td>
<td>CE 2</td>
<td>Age (35–49, 50–59, 60–69 years), gender</td>
<td>Age (35–49, 50–59, 60–69 years), gender</td>
<td></td>
</tr>
<tr>
<td>Hamilton</td>
<td>USA</td>
<td>Reductase inhibitors to prevent coronary heart disease</td>
<td>CE 2</td>
<td>Age, gender, risk profile</td>
<td>Age, gender, risk profile</td>
<td></td>
</tr>
<tr>
<td>Hillner</td>
<td>USA</td>
<td>Androgen blockade with flutamide in M1 prostate cancer</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hunink</td>
<td>USA</td>
<td>Computed tomography</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johannesson</td>
<td>Sweden</td>
<td>Preventive programme for hypertension</td>
<td>CE 2</td>
<td>Age (55, 70 years), gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kupersmith</td>
<td>USA</td>
<td>Implantable cardioverter–defibrillator for ventricular arrhythmias</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lindfors</td>
<td>USA</td>
<td>Mammography screening strategies</td>
<td>CE 2</td>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lindholm</td>
<td>Sweden</td>
<td>Treatment of hypertension</td>
<td>CE 2</td>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mangtani</td>
<td>UK</td>
<td>Hepatitis B vaccination</td>
<td>CE 2</td>
<td>High risk versus universal</td>
<td>High risk versus universal</td>
<td></td>
</tr>
<tr>
<td>Margolis</td>
<td>USA</td>
<td>Immunisation to prevent hepatitis B</td>
<td>CE 2</td>
<td>Age (perinatal, infant, adolescent)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mark</td>
<td>USA</td>
<td>Therapy for myocardial infarction</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>First author</th>
<th>Country</th>
<th>Type of intervention</th>
<th>Study type</th>
<th>Outcome measure</th>
<th>Effects in different subpopulations</th>
<th>Comparison of CER by subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midgette</td>
<td>USA</td>
<td>High-dose dexamethasone suppression testing versus inferior petrosal sinus sampling in the differential diagnosis of adrenocorticotrophin-dependent Cushing’s syndrome</td>
<td>CE</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plans</td>
<td>Spain</td>
<td>Smoking-cessation strategies</td>
<td>CE</td>
<td>2</td>
<td>Age</td>
<td>Age</td>
</tr>
<tr>
<td>Smith</td>
<td>USA</td>
<td>Vinorelbine, cisplatin, vindesine</td>
<td>CE</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bergquist</td>
<td>USA</td>
<td>Liver biopsy in rheumatoid arthritis patients treated with methotrexate</td>
<td>CE</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daya</td>
<td>Canada</td>
<td>Techniques for in vitro fertilisation</td>
<td>CE</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dusheiko</td>
<td>UK</td>
<td>Interferon-α for hepatitis B and C</td>
<td>CE</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kind</td>
<td>UK</td>
<td>SSRIs for depression</td>
<td>CE</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Law</td>
<td>USA</td>
<td>Glipizide versus glyburide</td>
<td>CE</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O’Neill</td>
<td>UK</td>
<td>Adjunctive therapy in epilepsy</td>
<td>CE</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Van Hout</td>
<td>Netherlands</td>
<td>C7E3 in percutaneous transluminal coronary angioplasty patients</td>
<td>CE</td>
<td>3</td>
<td>Weight, previous history of disease</td>
<td></td>
</tr>
<tr>
<td>Arrigoni</td>
<td>Italy</td>
<td>Diagnostic protocol for colonic neoplasms</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Babazono</td>
<td>Japan</td>
<td>Screening for gastric cancer</td>
<td>CE</td>
<td>4</td>
<td>Age (40–70 years), gender</td>
<td>Age (40–70 years), gender</td>
</tr>
<tr>
<td>Belinson</td>
<td>USA</td>
<td>Screening for women at increased risk of breast cancer</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bryan</td>
<td>UK</td>
<td>Mammography screening</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cuckle</td>
<td>UK</td>
<td>Antenatal screening for cystic fibrosis</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gronvald</td>
<td>Denmark</td>
<td>Diagnosis in women with menstrual disorders</td>
<td>CE</td>
<td>4</td>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>Gustafsson</td>
<td>Sweden</td>
<td>Prostate cancer screening</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kantrowitz</td>
<td>USA</td>
<td>Prostate cancer screening</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bergmann</td>
<td>France</td>
<td>Drug treatment for reflux oesophagitis</td>
<td>CE</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fendrick</td>
<td>USA</td>
<td>Diagnostic and treatment strategies for suspected peptic ulcer</td>
<td>CE</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sewett</td>
<td>Canada</td>
<td>Urinary stone disease management</td>
<td>CE</td>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>First author</th>
<th>Country</th>
<th>Type of intervention</th>
<th>Study type</th>
<th>Outcome measure†</th>
<th>Effects in different subpopulations</th>
<th>Comparison of CER by subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adhikari</td>
<td>Canada</td>
<td>Screening for tuberculosis</td>
<td>CE 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ferraz</td>
<td>Canada</td>
<td>Urate-lowering drugs in non-tophaceous recurrent gouty arthritis</td>
<td>CE 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glotzer</td>
<td>USA</td>
<td>Management of childhood lead poisoning</td>
<td>CE 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grina</td>
<td>Spain</td>
<td>Vaccination for hepatitis B</td>
<td>CE 6</td>
<td></td>
<td>Age (adults, children)</td>
<td>Age (adults, children)</td>
</tr>
<tr>
<td>Guyatt</td>
<td>UK</td>
<td>Control of <em>Ascaris</em> infection by chemotherapy</td>
<td>CE 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lieberman</td>
<td>USA</td>
<td>Colon cancer screening</td>
<td>CE 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lieu</td>
<td>USA</td>
<td>Varicella testing</td>
<td>CE 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Menzin</td>
<td>USA</td>
<td>Prevention of deep vein thrombosis in hip replacement</td>
<td>CE 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Owens</td>
<td>USA</td>
<td>Screening of surgeons for HIV</td>
<td>CE 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rouse</td>
<td>USA</td>
<td>Screening and treatment of asymptomatic bacteriuria in pregnancy to prevent pyelonephritis</td>
<td>CE 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Van Damme</td>
<td>Europe</td>
<td>Hepatitis B vaccination</td>
<td>CE 6</td>
<td>Age (adolescent, neonatal)</td>
<td>Age (adolescent, neonatal)</td>
<td></td>
</tr>
</tbody>
</table>

1996

<p>| Barr        | Canada  | Bone marrow transplantation | CE 1       |                  |                                     |                               |
| Benett      | USA     | Flutamide therapy in prostate cancer | CE 1       |                  |                                     |                               |
| Chang       | USA     | Hip arthroplasty | CE 1       | Age              | Age                                |                               |
| Danese      | USA     | Screening for mild thyroid failure | CE 1       | Gender           | Gender                             |                               |
| Derdeyn     | USA     | Screening for carotid atherosclerotic disease | CE 1       |                  |                                     |                               |
| Fiscella    | USA     | Nicotine patch for smoking cessation | CE 1       | Age (40, 60 years), gender | Age, gender                         |                               |
| Govert      | USA     | Collecting cytologic specimens in fibre-optic bronchoscopy | CE 1       |                  |                                     |                               |
| Gregor      | Canada  | Endoscopic retrograde cholangiopancreatography after idiopathic pancreatitis | CE 1       |                  |                                     |                               |
| Holtgrave   | USA     | Strategies for the prevention of HIV | CE 1       |                  |                                     |                               |
| Jonsson     | Sweden  | Fracture prevention in osteoporosis | CE 1       |                  |                                     |                               |</p>
<table>
<thead>
<tr>
<th>First author</th>
<th>Country</th>
<th>Type of intervention</th>
<th>Study type</th>
<th>Outcome measure</th>
<th>Effects in different subpopulations</th>
<th>Comparison of CER by subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kattan</td>
<td>USA</td>
<td>Interferon-α and conventional chemotherapy in leukaemia</td>
<td>CE</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kuntz</td>
<td>USA</td>
<td>Routine coronary angiography after acute myocardial infarction</td>
<td>CE</td>
<td>1</td>
<td>Age, gender</td>
<td>Age, gender</td>
</tr>
<tr>
<td>Kuntz</td>
<td>USA</td>
<td>Carotid endarterectomy</td>
<td>CE</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Livartowski</td>
<td>France</td>
<td>Vaccination against <em>Haemophilus influenzae</em></td>
<td>CE</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malter</td>
<td>USA</td>
<td>Lumbar discectomy</td>
<td>CE</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nichol</td>
<td>Canada</td>
<td>Emergency medical services in cardiac arrest</td>
<td>CE</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norum</td>
<td>Norway</td>
<td>Treatment of Hodgkin’s disease</td>
<td>CE</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Owens</td>
<td>USA</td>
<td>HIV screening</td>
<td>CE</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plans</td>
<td>Spain</td>
<td>Breast cancer detection</td>
<td>CE</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Richards</td>
<td>UK</td>
<td>Home parenteral nutrition</td>
<td>CE</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sarasin</td>
<td>Switzerland</td>
<td>Prophylactic anticoagulation prolonged after hospital discharge following general surgery</td>
<td>CE</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sculpher</td>
<td>UK</td>
<td>Laser-assisted angioplasty for peripheral arterial occlusion</td>
<td>CE</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>St Leger</td>
<td>UK</td>
<td>Screening for abdominal aortic aneurysm</td>
<td>CE</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes Control and Complications Trials Group</td>
<td>USA</td>
<td>Diabetes management</td>
<td>CE</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>van Hout</td>
<td>Netherlands</td>
<td>Stent implantation versus percutaneous transluminal coronary angioplasty</td>
<td>CE</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wyatt</td>
<td>USA</td>
<td>Multichannel cochlear implant</td>
<td>CE</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhang</td>
<td>USA</td>
<td>Clozapine monitoring</td>
<td>CE</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ashraf</td>
<td>USA</td>
<td>Pravastatin in the prevention of coronary heart disease</td>
<td>CE</td>
<td>2</td>
<td>Risk profile</td>
<td>Risk profile</td>
</tr>
<tr>
<td>Beutels</td>
<td>Germany</td>
<td>Vaccination for varicella</td>
<td>CE</td>
<td>2</td>
<td>Age</td>
<td>Age</td>
</tr>
<tr>
<td>Chertow</td>
<td>USA</td>
<td>Cancer screening in end-stage renal disease</td>
<td>CE</td>
<td>2</td>
<td>Age (20–85 years), ethnicity (black, white)</td>
<td>Age (20–85 years), ethnicity (black, white)</td>
</tr>
<tr>
<td>Chesebro</td>
<td>USA</td>
<td>Colposcopy</td>
<td>CE</td>
<td>2</td>
<td>Disease severity</td>
<td>Disease severity</td>
</tr>
</tbody>
</table>

*Continued*
<table>
<thead>
<tr>
<th>First author</th>
<th>Country</th>
<th>Type of intervention</th>
<th>Study type</th>
<th>Outcome measure¹</th>
<th>Effects in different subpopulations</th>
<th>Comparison of CER by subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covens</td>
<td>Canada</td>
<td>Paclitaxel and cisplatin in ovarian carcinoma</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daly</td>
<td>UK</td>
<td>Hormone replacement therapy</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evans</td>
<td>Canada</td>
<td>Support care in lung cancer</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evans</td>
<td>Canada</td>
<td>Gemcitabine in lung cancer</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freedberg</td>
<td>USA</td>
<td>Secondary prophylaxis in AIDS</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girgis</td>
<td>Australia</td>
<td>Screening for melanoma</td>
<td>CE 2</td>
<td>Gender</td>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Hillner</td>
<td>USA</td>
<td>Stereotaxic biopsy for breast abnormalities</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johannesson</td>
<td>Sweden</td>
<td>Hypertension treatment</td>
<td>CE 2</td>
<td>Age, gender</td>
<td>Age, gender</td>
<td></td>
</tr>
<tr>
<td>Johannesson</td>
<td>Sweden</td>
<td>Lipid-lowering therapies</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jonsson</td>
<td>Sweden</td>
<td>Cholesterol lowering with simvastatin</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Langham</td>
<td>UK</td>
<td>Health checks by nurses in the Oxcheck study</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lindholm</td>
<td>Sweden</td>
<td>Community-based cardiovascular prevention programme</td>
<td>CE 2</td>
<td>Age (30, 40, 50, 60 years), gender, risk profile</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Messori</td>
<td>Italy</td>
<td>Adjuvant chemotherapy in breast cancer</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Messori</td>
<td>Italy</td>
<td>Paclitaxel in ovarian cancer</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Messori</td>
<td>Italy</td>
<td>Adjuvant chemotherapy in colorectal cancer</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meythaler</td>
<td>USA</td>
<td>Screening for brain injury</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Michel</td>
<td>Netherlands</td>
<td>Captopril in left ventricular dysfunction</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parsonnet</td>
<td>USA</td>
<td>Helicobacter pylori screening</td>
<td>CE 2</td>
<td>Age, ethnic group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pharoah</td>
<td>UK</td>
<td>Cholesterol-lowering therapies</td>
<td>CE 2</td>
<td>Age, gender</td>
<td>Age, gender</td>
<td></td>
</tr>
<tr>
<td>Salkeld</td>
<td>Australia</td>
<td>Screening for colorectal cancer with faecal occult blood</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sarasin</td>
<td>Switzerland</td>
<td>Screening for hepatocellular carcinoma</td>
<td>CE 2</td>
<td>Risk profile</td>
<td>Risk profile</td>
<td></td>
</tr>
<tr>
<td>Schechter</td>
<td>USA</td>
<td>Rescreening by Papnet testing</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shield</td>
<td>USA</td>
<td>Cadaveric kidney transplantation</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>First author</th>
<th>Country</th>
<th>Type of intervention</th>
<th>Study type</th>
<th>Outcome measure¹</th>
<th>Effects in different subpopulations</th>
<th>Comparison of CER by subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smales</td>
<td>Australia</td>
<td>Dental restorative materials used in various classes of cavity preparations</td>
<td>CE 2</td>
<td></td>
<td>Age, risk profile</td>
<td>Age, risk profile</td>
</tr>
<tr>
<td>Tosteson</td>
<td>USA</td>
<td>Coronary care unit versus intermediate care in chest pain patients</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>van Erkel</td>
<td>Netherlands</td>
<td>Spiral computed tomography angiography</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Villari</td>
<td>Italy</td>
<td>HIV testing in intravenous drug users</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Warner</td>
<td>USA</td>
<td>Smoking cessation</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waugh</td>
<td>UK</td>
<td>Cervical cancer screening</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waugh</td>
<td>UK</td>
<td>Cervical cancer screening</td>
<td>CE 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wonderling</td>
<td>UK</td>
<td>Cardiovascular screening</td>
<td>CE 2</td>
<td>Gender</td>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Blaufox</td>
<td>USA</td>
<td>Diagnosis and therapy of renovascular hypertension</td>
<td>CE 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chandhoke</td>
<td>USA</td>
<td>Treatment options for staghorn calculi</td>
<td>CE 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goldfarb</td>
<td>USA</td>
<td>In vitro fertilisation</td>
<td>CE 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Granberg</td>
<td>Sweden</td>
<td>Intracytoplasmic sperm injection</td>
<td>CE 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palmer</td>
<td>Italy</td>
<td>Cisplatin in lung cancer</td>
<td>CE 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buck</td>
<td>UK</td>
<td>Health-promotion strategies against coronary heart disease</td>
<td>CE 4</td>
<td>Gender, age (adults)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friedland</td>
<td>USA</td>
<td>High-risk register and auditory brainstem response</td>
<td>CE 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gill</td>
<td>UK</td>
<td>Screening for diabetes in chiropody clinics</td>
<td>CE 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hillner</td>
<td>USA</td>
<td>Three regimens using Vinorelbine for non-small-cell lung cancer</td>
<td>CE 4</td>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manus</td>
<td>Netherlands</td>
<td>Screening for gastrointestinal cancer</td>
<td>CE 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plans</td>
<td>Spain</td>
<td>Breast cancer screening</td>
<td>CE 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rouse</td>
<td>USA</td>
<td>Management of presumed susceptible varicella in (chickenpox) exposed gravida</td>
<td>CE 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vergnon</td>
<td>France</td>
<td>Screening of donated blood</td>
<td>CE 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wright</td>
<td>UK</td>
<td>Treatment of Barrett’s cancer</td>
<td>CE 4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Continued*
<table>
<thead>
<tr>
<th>First author</th>
<th>Country</th>
<th>Type of intervention</th>
<th>Study type</th>
<th>Outcome measure</th>
<th>Effects in different subpopulations</th>
<th>Comparison of CER by subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anis</td>
<td>Canada</td>
<td>Cyclosporine in rheumatoid arthritis</td>
<td>CE 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cieslak</td>
<td>USA</td>
<td>Granisetron in emesis in paediatry</td>
<td>CE 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geitung</td>
<td>Norway</td>
<td>Ultrasonographic screening for congenital hip dysplasia in newborns</td>
<td>CE 5</td>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genc</td>
<td>Sweden</td>
<td>Screening for Chlamydia</td>
<td>CE 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jonsson</td>
<td>Sweden</td>
<td>Helicobacter pylori eradication</td>
<td>CE 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kinlay</td>
<td>Australia</td>
<td>Coronary angioplasty versus medical treatment</td>
<td>CE 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schectman</td>
<td>USA</td>
<td>Cholesterol-lowering therapies</td>
<td>CE 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Siggaard</td>
<td>Denmark</td>
<td>Weight-loss strategies</td>
<td>CE 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>van Doorslaer</td>
<td>Germany</td>
<td>Onychomycosis treatment</td>
<td>CE 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vergnenegre</td>
<td>France</td>
<td>Chemotherapy in lung cancer</td>
<td>CE 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ankjaer-Jansen</td>
<td>Denmark</td>
<td>Pharmaceutical treatments in osteoporosis</td>
<td>CE 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ben-Menachem</td>
<td>USA</td>
<td>Prophylaxis for gastrointestinal haemorrhage</td>
<td>CE 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boivin</td>
<td>Canada</td>
<td>Screening for glaucoma</td>
<td>CE 6</td>
<td>Age (40–79, 65–79 years)</td>
<td>Age (40–79, 65–79 years)</td>
<td></td>
</tr>
<tr>
<td>Bos</td>
<td>Netherlands</td>
<td>Use of collagen haemostatic closure device</td>
<td>CE 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connely</td>
<td>Australia</td>
<td>Nutrition support in Wernicke–Korsakoff syndrome</td>
<td>CE 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fernandez de Gatta</td>
<td>Spain</td>
<td>Vancomycin monitoring</td>
<td>CE 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ecker</td>
<td>USA</td>
<td>Screening for HIV in pregnancy</td>
<td>CE 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gibbs</td>
<td>UK</td>
<td>Antiemetic use of Granisetron</td>
<td>CE 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kahn</td>
<td>USA</td>
<td>HIV prevention targeting</td>
<td>CE 6</td>
<td>Risk profile</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Messori</td>
<td>Italy</td>
<td>Pharmacoeconomic considerations in lung cancer</td>
<td>CE 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Michel</td>
<td>France</td>
<td>Treatments for hepatitis C</td>
<td>CE 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miller</td>
<td>USA</td>
<td>Poliovirus vaccine in routine childhood immunisation</td>
<td>CE 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randolph</td>
<td>USA</td>
<td>Acyclovir to prevent neonatal herpes</td>
<td>CE 6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>First author</th>
<th>Country</th>
<th>Type of intervention</th>
<th>Study type</th>
<th>Outcome measure</th>
<th>Effects in different subpopulations</th>
<th>Comparison of CER by subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rouse</td>
<td>USA</td>
<td>Elective caesarean delivery for fetal macrosoma</td>
<td>CE 6</td>
<td>Risk profile</td>
<td>Risk profile</td>
<td></td>
</tr>
<tr>
<td>Smith</td>
<td>USA</td>
<td>Treatment of haemophilia A</td>
<td>CE 6</td>
<td>Risk profile</td>
<td>Risk profile</td>
<td></td>
</tr>
<tr>
<td>Tang</td>
<td>USA</td>
<td>Ondansetron and Droperidol antiemetic therapy</td>
<td>CE 6</td>
<td>Risk profile</td>
<td>Risk profile</td>
<td></td>
</tr>
<tr>
<td>Vick</td>
<td>UK</td>
<td>Antenatal anti-D prophylaxis</td>
<td>CE 6</td>
<td>Risk profile</td>
<td>Risk profile</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AuBuchon</td>
<td>USA</td>
<td>Expanded HIV-testing protocols for donated blood</td>
<td>CE 1</td>
<td>Risk profile</td>
<td>Risk profile</td>
<td></td>
</tr>
<tr>
<td>Barkun</td>
<td>Canada</td>
<td>Extracorporeal gall-bladder stone shock wave lithotripsy versus laparoscopic cholecystectomy</td>
<td>CE 1</td>
<td>Age, gender</td>
<td>Age, gender</td>
<td></td>
</tr>
<tr>
<td>Barry</td>
<td>USA</td>
<td>Evaluation of office-based strategies of suspected urinary tract infections in ambulatory care</td>
<td>CE 1</td>
<td>Risk profile</td>
<td>Risk profile</td>
<td></td>
</tr>
<tr>
<td>Bennett</td>
<td>USA</td>
<td>Single course of interferon-α-2b in patients with histologically mild chronic hepatitis C</td>
<td>CE 1</td>
<td>Age (20–70 years in 5-year intervals), risk profile</td>
<td>Age (20–70 years in 5-year intervals), risk profile</td>
<td></td>
</tr>
<tr>
<td>Brasel</td>
<td>USA</td>
<td>Treatment of contaminated appendectomy wounds</td>
<td>CE 1</td>
<td>Risk profile</td>
<td>Risk profile</td>
<td></td>
</tr>
<tr>
<td>Busch</td>
<td>USA</td>
<td>Screening blood donors for antibody to hepatitis B core antigen as a way of detecting window-phase HIV type 1 infections</td>
<td>CE 1</td>
<td>Gender, ethnicity (white, black, Hispanic, other), age (&lt;25, 26–35, 36–45, &gt;45 years)</td>
<td>Gender, ethnicity (white, black, Hispanic, other), age (&lt;25, 26–35, 36–45, &gt;45 years)</td>
<td></td>
</tr>
<tr>
<td>Chouinard</td>
<td>Canada</td>
<td>Schizophrenic patients treated with risperidone or haloperidol</td>
<td>CE 1</td>
<td>Disease severity (mild, moderate, severe), age, gender mix, age of schizophrenia onset, age at first hospitalisation</td>
<td>Gender, ethnicity (white, black, Hispanic, other), age (&lt;25, 26–35, 36–45, &gt;45 years)</td>
<td></td>
</tr>
<tr>
<td>Cleland</td>
<td>UK</td>
<td>Treatment of angina</td>
<td>CE 1</td>
<td>Risk profile</td>
<td>Risk profile</td>
<td></td>
</tr>
<tr>
<td>Coley</td>
<td>USA</td>
<td>Early detection of prostate cancer</td>
<td>CE 1</td>
<td>Age (55, 65, 75 years)</td>
<td>Age (55, 65, 75 years)</td>
<td></td>
</tr>
<tr>
<td>Cromwell</td>
<td>USA</td>
<td>Guidelines for smoking cessation</td>
<td>CE 1</td>
<td>Risk profile</td>
<td>Risk profile</td>
<td></td>
</tr>
<tr>
<td>Cronan</td>
<td>USA</td>
<td>Social support and education interventions</td>
<td>CE 1</td>
<td>Risk profile</td>
<td>Risk profile</td>
<td></td>
</tr>
<tr>
<td>Cronenwett</td>
<td>USA</td>
<td>Carotid endarterectomy in asymptomatic patients</td>
<td>CE 1</td>
<td>Age at presentation (60–80 years)</td>
<td>Age at presentation (60–80 years)</td>
<td></td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>First author</th>
<th>Country</th>
<th>Type of intervention</th>
<th>Study type</th>
<th>Outcome measure (^1)</th>
<th>Effects in different subpopulations</th>
<th>Comparison of CER by subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastman</td>
<td>USA</td>
<td>Treatments for non-insulin-dependent diabetes mellitus</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ebell</td>
<td>USA</td>
<td>Treatment of dyspepsia</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freedberg</td>
<td>USA</td>
<td>Prophylaxis for disseminated <em>Mycobacterium avium</em> complex infection in patients with AIDS</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gartner</td>
<td>USA</td>
<td>Lung transplantation</td>
<td>CE</td>
<td>I</td>
<td>Age (10–64, &lt;64, &gt;64 years)</td>
<td>Age (10–64, &lt;64, &gt;64 years)</td>
</tr>
<tr>
<td>Graham</td>
<td>USA</td>
<td>Airbags by seat position</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hakama</td>
<td>Finland</td>
<td>Screening in the Nordic countries up to the year 2017</td>
<td>CE</td>
<td>I</td>
<td>Age, type of cancer</td>
<td>Age, type of cancer</td>
</tr>
<tr>
<td>Hamek</td>
<td>USA</td>
<td>Dialysis and continuing aggressive care in seriously ill hospitalised adults</td>
<td>CE</td>
<td>I</td>
<td>Survival probability</td>
<td>Survival probability</td>
</tr>
<tr>
<td>Harris</td>
<td>USA</td>
<td>Proton pump inhibitors or histamine-2 receptor antagonists for the prevention of recurrences of erosive reflux oesophagitis</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harris</td>
<td>USA</td>
<td>Prevention of recurrences of erosive reflux oesophagitis analysis of maintenance proton pump inhibitors</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Henschke</td>
<td>USA</td>
<td>Algorithms in the diagnosis of pulmonary embolism</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hillner</td>
<td>USA</td>
<td>Adjuvant interferon-α-2b in high-risk melanoma</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holtgrave</td>
<td>USA</td>
<td>HIV-prevention programmes</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hornberger</td>
<td>USA</td>
<td>Repeat medical procedures (kidney transplantation)</td>
<td>CE</td>
<td>I</td>
<td>Age (0–14, 15–24, 25–34, 35–44, 45–54, ≥55 years)</td>
<td>Age (0–14, 15–24, 25–34, 35–44, 45–54, ≥55 years)</td>
</tr>
<tr>
<td>Huber</td>
<td>USA</td>
<td>Intraoperative autologous transfusion during elective infrarenal aortic reconstruction</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johannesson</td>
<td>Sweden</td>
<td>Treatment of hypertension</td>
<td>CE</td>
<td>I</td>
<td>Gender, age (&lt;45, 45–69, &gt;70 years)</td>
<td>Gender, age (&lt;45, 45–69, &gt;70 years)</td>
</tr>
<tr>
<td>Kallmes</td>
<td>USA</td>
<td>Routine angiography after surgery for ruptured intracranial aneurysms</td>
<td>CE</td>
<td>I</td>
<td>Age</td>
<td>Age</td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>First author</th>
<th>Country</th>
<th>Type of intervention</th>
<th>Study type</th>
<th>Outcome measure</th>
<th>Effects in different subpopulations</th>
<th>Comparison of CER by subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kim</td>
<td>USA</td>
<td>6 and 12 months of interferon-α therapy for hepatitis C</td>
<td>CE</td>
<td>I</td>
<td>Age (30, 40, 50, 60 years)</td>
<td>Age (30, 40, 50, 60 years)</td>
</tr>
<tr>
<td>King</td>
<td>USA</td>
<td>Management of pituitary microadenomas</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>King</td>
<td>USA</td>
<td>Anterior temporal lobectomy for intractable temporal lobe epilepsy</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Konski</td>
<td>USA</td>
<td>Malignant glioma treatment protocol</td>
<td>CE</td>
<td>I</td>
<td>Age (&lt;50, &gt;50 years)</td>
<td>Age (&lt;50, &gt;50 years)</td>
</tr>
<tr>
<td>Langfit</td>
<td>USA</td>
<td>Anterotemporal lobectomy in medically intractable complex partial epilepsy</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liberato</td>
<td>Italy</td>
<td>Interferon-α in chronic myelogenous leukaemia</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liljegren</td>
<td>Sweden</td>
<td>Routine postoperative radiotherapy after sector resection and axillary dissection for breast cancer stage I</td>
<td>CE</td>
<td>I</td>
<td>Low risk versus high risk (age related)</td>
<td>Low risk versus high risk (age related)</td>
</tr>
<tr>
<td>McElwee</td>
<td>USA</td>
<td>Glycoprotein IIB and IIIA inhibitors in patients with acute ischaemic coronary syndromes</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>McMurray</td>
<td>UK</td>
<td>ACE inhibitor treatments post-myocardial infarction</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>McNamara</td>
<td>USA</td>
<td>Echocardiographic identification of cardiovascular sources of emboli to guide clinical management of stroke</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Messori</td>
<td>Italy</td>
<td>Analogous bone marrow transplantation in patients with relapsed non-Hodgkin’s lymphoma</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Messori</td>
<td>Italy</td>
<td>Advanced HIV infection treated with zidovudine monotherapy</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Messori</td>
<td>Italy</td>
<td>Interferon as adjuvant therapy in high-risk resected cutaneous melanoma</td>
<td>CE</td>
<td>I</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>First author</th>
<th>Country</th>
<th>Type of intervention</th>
<th>Study type</th>
<th>Outcome measure</th>
<th>Effects in different subpopulations</th>
<th>Comparison of CER by subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mooney USA</td>
<td>Life-long screening of patients with intermediate thickness cutaneous melanoma for asymptomatic pulmonary recurrences</td>
<td>CE 1</td>
<td>I</td>
<td>Age at diagnosis, gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moore USA</td>
<td>Prophylactic treatment with oral ganciclovir for cytomegalovirus retinitis</td>
<td>CE 1</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mushlin USA</td>
<td>Magnetic resonance imaging for patients with equivocal neurological symptoms</td>
<td>CE 1</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norum Norway</td>
<td>Lumpectomy or mastectomy</td>
<td>CE 1</td>
<td>I</td>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norum Norway</td>
<td>Adjuvant chemotherapy in Duke B and C colorectal carcinoma</td>
<td>CE 1</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obuchowski USA</td>
<td>Non-invasive screening for patients with asymptomatic neck bruits</td>
<td>CE 1</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ortega Canada</td>
<td>Paclitaxel in combination with cisplatin for patients with advanced ovarian cancer</td>
<td>CE 1</td>
<td>I</td>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Owens USA</td>
<td>Implantable cardioverter-defibrillators relative to amidarone for prevention of sudden cardiac death</td>
<td>CE 1</td>
<td>I</td>
<td>Age, health risk</td>
<td>Age, health risk</td>
<td></td>
</tr>
<tr>
<td>Pinkerton USA</td>
<td>Chemoprophylaxis after occupational exposure to HIV</td>
<td>CE 1</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Porter Canada</td>
<td>Surgery versus stereotactic radiosurgery for small operable cerebral arteriovenous malformations</td>
<td>CE 1</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Revicki USA</td>
<td>Newer antidepressants compared with tricyclic antidepressants in managed-care settings</td>
<td>CE 1</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smith USA</td>
<td>Hepatitis A vaccination in healthcare workers</td>
<td>CE 1</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Victorian Infant Collaborative Study Group</td>
<td>Exogenous surfactant</td>
<td>CE 1</td>
<td>I</td>
<td>Birthweight subgroups</td>
<td>Birthweight subgroups</td>
<td></td>
</tr>
<tr>
<td>Yee USA, UK</td>
<td>Taxane therapy</td>
<td>CE 1</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First author</td>
<td>Country</td>
<td>Type of intervention</td>
<td>Study type</td>
<td>Outcome measure&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Effects in different subpopulations</td>
<td>Comparison of CER by subgroups</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>------------</td>
<td>------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Barry</td>
<td>Ireland</td>
<td>Treatment of ruptured abdominal aortic aneurysm</td>
<td>CE</td>
<td>2</td>
<td>Gender, age (&lt;70, &gt;70 years)</td>
<td>Gender, age (&lt;70, &gt;70 years)</td>
</tr>
<tr>
<td>Durand-Zaleski</td>
<td>France</td>
<td>Commercial bags or glass bottles for total parenteral nutrition</td>
<td>CE</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lieu</td>
<td>USA</td>
<td>Primary angioplasty for acute myocardial infarction</td>
<td>CE</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lorenzoni</td>
<td>Italy</td>
<td>tPA versus streptokinase in acute myocardial infarction</td>
<td>CE</td>
<td>2</td>
<td>Age (&lt;75, &gt;75 years), disease type</td>
<td>Age (&lt;75, &gt;75 years), disease type</td>
</tr>
<tr>
<td>Vanara</td>
<td>Italy</td>
<td>Breast cancer screening</td>
<td>CE</td>
<td>2</td>
<td>Disease type</td>
<td></td>
</tr>
<tr>
<td>Barnett</td>
<td>USA</td>
<td>Inpatient substance-abuse treatment</td>
<td>CE</td>
<td>3</td>
<td>Age, marital status, income, drug use, medical conditions</td>
<td></td>
</tr>
<tr>
<td>Collins</td>
<td>Canada</td>
<td>Infertility diagnosis and treatment</td>
<td>CE</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>de Lissovoy</td>
<td>USA</td>
<td>Long-term intrathecal morphine therapy for pain associated with back surgery syndrome</td>
<td>CE</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kraft</td>
<td>USA</td>
<td>Supplementary services during methadone maintenance</td>
<td>CE</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mol</td>
<td>Netherlands</td>
<td>Salpingectomy versus conservative tubal surgery in women with tubal pregnancy</td>
<td>CE</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sykes</td>
<td>UK</td>
<td>Ondansetron versus a chlorpromazine + dexamethasone combination for the prevention of nausea and vomiting</td>
<td>CE</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Van Voorhis</td>
<td>USA</td>
<td>Infertility treatments</td>
<td>CE</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Woods</td>
<td>USA</td>
<td>Antidepressant treatment</td>
<td>CE</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdool-Carrim</td>
<td>South Africa</td>
<td>Prophylaxis against deep vein thrombosis in elective hip replacement</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ades</td>
<td>USA</td>
<td>Cardiac rehabilitation after myocardial infarction</td>
<td>CE</td>
<td>4</td>
<td>Risk profile (several characteristics)</td>
<td>Risk profile (several characteristics)</td>
</tr>
<tr>
<td>Barosi</td>
<td>Italy</td>
<td>Occult cancer in patients with idiopathic deep vein thrombosis</td>
<td>CE</td>
<td>4</td>
<td>Gender, age</td>
<td>Age (60–69 years)</td>
</tr>
<tr>
<td>Baxter</td>
<td>UK</td>
<td>Community-based heart-health-promotion project in England</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>First author</th>
<th>Country</th>
<th>Type of intervention</th>
<th>Study type</th>
<th>Outcome measure</th>
<th>Effects in different subpopulations</th>
<th>Comparison of CER by subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caro</td>
<td>UK</td>
<td>Primary coronary prevention with pravastatin</td>
<td>CE</td>
<td>4</td>
<td>Risk profile (several characteristics)</td>
<td></td>
</tr>
<tr>
<td>Castillo</td>
<td>USA</td>
<td>Thrombolytic therapy for acute myocardial infarction</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crogan</td>
<td>USA</td>
<td>Treating nicotine dependence</td>
<td>CE</td>
<td>4</td>
<td>Gender, risk profile</td>
<td></td>
</tr>
<tr>
<td>Durand-Zaleski</td>
<td>USA, France</td>
<td>Hepatic arterial infusion chemotherapy in treatment of non-resectable colorectal liver metastases</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earle</td>
<td>Canada</td>
<td>Support care in stage IV non-small-cell lung cancer</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elit</td>
<td>Canada</td>
<td>Paclitaxel as first-line therapy for advanced ovarian cancer</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evans</td>
<td>Canada</td>
<td>Gemcitabine in stage IV non-small-cell lung cancer</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evans</td>
<td>Canada</td>
<td>Interventions for stage III non-small-cell lung cancer</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gearhart</td>
<td>USA</td>
<td>Helicopter emergency service</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hlatky</td>
<td>USA</td>
<td>Coronary angioplasty versus coronary bypass surgery</td>
<td>CE</td>
<td>4</td>
<td>Age, gender, education, marital status</td>
<td></td>
</tr>
<tr>
<td>Johannesson</td>
<td>Scandinavia</td>
<td>Simvastatin treatment to lower cholesterol levels</td>
<td>CE</td>
<td>4</td>
<td>Age (35, 59, 70 years), gender</td>
<td>Age (35, 59, 70 years), gender</td>
</tr>
<tr>
<td>Louis-Jacques</td>
<td>USA</td>
<td>Interferon therapy in children with chronic active hepatitis B</td>
<td>CE</td>
<td>4</td>
<td>Age (2, 12, 25 years)</td>
<td>Age (2, 12, 25 years)</td>
</tr>
<tr>
<td>Mandelblatt</td>
<td>USA</td>
<td>Cervical and breast cancer screening in public hospital emergency room</td>
<td>CE</td>
<td>4</td>
<td>Age (&lt;40 to &gt;65 years), race (black, Hispanic, white, other), education (&lt;10 to &gt;12 years)</td>
<td></td>
</tr>
<tr>
<td>Matsunaga</td>
<td>Japan</td>
<td>Mass screening for cervical cancer</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>McGuire</td>
<td>USA</td>
<td>Paclitaxel as alternative combination therapy for advanced ovarian cancer</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mehta</td>
<td>USA</td>
<td>Radiosurgery versus resection for single brain metastases</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morris</td>
<td>UK</td>
<td>Cholesterol-modifying pharmacotherapy</td>
<td>CE</td>
<td>4</td>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>Munro</td>
<td>UK</td>
<td>Physical activity at age &gt;65 years</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First author</td>
<td>Country</td>
<td>Type of intervention</td>
<td>Study type</td>
<td>Outcome measure</td>
<td>Effects in different subpopulations</td>
<td>Comparison of CER by subgroups</td>
</tr>
<tr>
<td>--------------</td>
<td>---------</td>
<td>----------------------</td>
<td>------------</td>
<td>-----------------</td>
<td>-------------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Nettleman</td>
<td>USA</td>
<td>Preventing tuberculosis in physicians using tuberculin skin testing or a hypothetical vaccine</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norum</td>
<td>Norway</td>
<td>Follow-up programme in colorectal cancer</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perreault</td>
<td>Canada</td>
<td>HMG-CoA reductase inhibitors and fibrates in different types of primary hyperlipidaemia</td>
<td>CE</td>
<td>4</td>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Perrier</td>
<td>Switzerland</td>
<td>Non-invasive diagnostic aids in suspected pulmonary embolism</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raab</td>
<td>USA</td>
<td>Cervical–vaginal rescreening</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raab</td>
<td>USA</td>
<td>Sputum cytology in the diagnosis of lung cancer</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratcliffe</td>
<td>Scotland</td>
<td>Mass-media-led antismoking campaign</td>
<td>CE</td>
<td>4</td>
<td>Age, gender, employment status</td>
<td></td>
</tr>
<tr>
<td>Riviere</td>
<td>Canada</td>
<td>Simvastatin in the secondary prevention of coronary artery disease</td>
<td>CE</td>
<td>4</td>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>Rose</td>
<td>USA</td>
<td>Cytomegalovirus disease prevention in patients with AIDS</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salzmann</td>
<td>USA</td>
<td>Screening mammography guidelines to include women aged 40–49 years</td>
<td>CE</td>
<td>4</td>
<td>Age (40–69 versus 50–69 years)</td>
<td>Age (40–69 versus 50–69 years)</td>
</tr>
<tr>
<td>Scharfstein</td>
<td>USA</td>
<td>Fluconazole prophylaxis against primary systemic fungal infections in AIDS patients</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secker-Walker</td>
<td>USA</td>
<td>Mass media programme to prevent smoking among adolescents</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simons</td>
<td>USA</td>
<td>Inpatient initiation of antiarrhythmic therapy for supraventricular tachycardia</td>
<td>CE</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tosteson</td>
<td>USA</td>
<td>Population-wide educational approaches to reduce serum cholesterol levels</td>
<td>CE</td>
<td>4</td>
<td>Age (35–84 years)</td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>USA</td>
<td>Protocol for ovarian cancer screening</td>
<td>CE</td>
<td>4</td>
<td>Age (50–80 years)</td>
<td></td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>First author</th>
<th>Country</th>
<th>Type of intervention</th>
<th>Study type</th>
<th>Outcome measure¹</th>
<th>Effects in different subpopulations</th>
<th>Comparison of CER by subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vale</td>
<td>UK</td>
<td>Thrombolysis in a remote rural community</td>
<td>CE 4</td>
<td></td>
<td>Age (35–39, 40–44, 45–49, 50–54, 55–59, 60–64, 65–69 years), gender</td>
<td></td>
</tr>
<tr>
<td>Warmerdam</td>
<td>Germany, Netherlands</td>
<td>Mammography screening</td>
<td>CE 4</td>
<td></td>
<td>Age (35–39, 40–44, 45–49, 50–54, 55–59, 60–64, 65–69 years), gender</td>
<td></td>
</tr>
<tr>
<td>Wasley</td>
<td>USA</td>
<td>Nicotine transdermal patch for smoking cessation</td>
<td>CE 4</td>
<td></td>
<td>Age (infants, 10 years, 12 years)</td>
<td>Age (infants, 10 years, 12 years)</td>
</tr>
<tr>
<td>Webb</td>
<td>USA</td>
<td>Treatments in advanced oesophagogastric cancer</td>
<td>CE 4</td>
<td></td>
<td>Age, gender</td>
<td></td>
</tr>
<tr>
<td>Wiebe</td>
<td>Canada</td>
<td>Hepatitis B immunisation in a low-incidence area</td>
<td>CE 4</td>
<td></td>
<td>Age (infants, 10 years, 12 years)</td>
<td>Age (infants, 10 years, 12 years)</td>
</tr>
<tr>
<td>Agboola</td>
<td>Canada</td>
<td>Routine follow-up after curative treatment for endometrial cancer</td>
<td>CE 5</td>
<td></td>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>Batchelder</td>
<td>USA</td>
<td>Routine dilated pupil screening</td>
<td>CE 5</td>
<td></td>
<td>Age (20–39, 40–64, ≥65 years)</td>
<td>Age (20–39, 40–64, ≥65 years)</td>
</tr>
<tr>
<td>Cameron</td>
<td>Australia</td>
<td>Questionnaire for the detection of unrecognised coronary heart disease</td>
<td>CE 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cecchini</td>
<td>Italy</td>
<td>Two modalities of management of women with cytologic evidence of squamous or glandular atypia</td>
<td>CE 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colin</td>
<td>France</td>
<td>Screening blood donors for hepatitis C and non-A, non-B, non-C hepatitis</td>
<td>CE 5</td>
<td></td>
<td>Age of donors, gender</td>
<td></td>
</tr>
<tr>
<td>Flannelly</td>
<td>UK</td>
<td>Immediate colposcopy or cytological surveillance for women with mild dyskaryosis</td>
<td>CE 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Francois</td>
<td>France</td>
<td>Detection of hearing impairment by transient evoked otoacoustic emissions</td>
<td>CE 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goldberg- Kahn</td>
<td>USA</td>
<td>Different strategies in the work-up of solitary radiographic lung lesions</td>
<td>CE 5</td>
<td></td>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>Kallmes</td>
<td>USA</td>
<td>Angiography performed during surgery for ruptured intracranial aneurysms</td>
<td>CE 5</td>
<td></td>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>Kaufman</td>
<td>USA</td>
<td>Human papillomavirus testing as triage for atypical squamous cells of undetermined significance and low-grade squamous intraepithelial lesions</td>
<td>CE 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First author</td>
<td>Country</td>
<td>Type of intervention</td>
<td>Study type</td>
<td>Outcome measure</td>
<td>Effects in different subpopulations</td>
<td>Comparison of CER by subgroups</td>
</tr>
<tr>
<td>--------------</td>
<td>---------</td>
<td>----------------------</td>
<td>------------</td>
<td>----------------</td>
<td>------------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Netser</td>
<td>USA</td>
<td>Examination of routine and non-routine tonsil and adenoid specimens</td>
<td>CE</td>
<td>5</td>
<td>Age, gender</td>
<td></td>
</tr>
<tr>
<td>Pollitt</td>
<td>UK</td>
<td>Neonatal screening</td>
<td>CE</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seradour</td>
<td>France</td>
<td>Breast cancer screening</td>
<td>CE</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shaw</td>
<td>USA</td>
<td>Screening filter test for urinary tract infection in children</td>
<td>CE</td>
<td>5</td>
<td>Age (&lt;2, &gt;2 years)</td>
<td></td>
</tr>
<tr>
<td>Siegman-Igra</td>
<td>USA</td>
<td>Diagnosis of vascular catheter-related bloodstream infection</td>
<td>CE</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuck</td>
<td>UK</td>
<td>Screening for primary open-angle glaucoma</td>
<td>CE</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bassett</td>
<td>Australia</td>
<td>Screening for haemochromatosis</td>
<td>CE</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bates</td>
<td>USA</td>
<td>Use of dexrazoxane in preventing anthracycline-induced cardiotoxicity in patients with stage IIIB or IV metastatic breast cancer</td>
<td>CE</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brasel</td>
<td>USA</td>
<td>Prevention of pulmonary embolus in high-risk trauma patients</td>
<td>CE</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Garton</td>
<td>Scotland</td>
<td>Perimenopausal bone density screening</td>
<td>CE</td>
<td>6</td>
<td>Age (45–54, 55–64, 65–74, 75–84, ≥85 years)</td>
<td></td>
</tr>
<tr>
<td>Garuz</td>
<td>Spain</td>
<td>Vaccination against hepatitis B virus</td>
<td>CE</td>
<td>6</td>
<td>Adolescents, infants and screening of pregnant women; acute versus chronic</td>
<td>Children, adolescents, combined, adolescents plus screening programme for pregnant women</td>
</tr>
<tr>
<td>Gray</td>
<td>UK</td>
<td>Varicella vaccination for healthcare workers</td>
<td>CE</td>
<td>6</td>
<td>Staff versus patients</td>
<td>Staff versus patients</td>
</tr>
<tr>
<td>Hankey</td>
<td>Australia</td>
<td>Treatment of people with reversible ischaemic attacks of the brain and eye</td>
<td>CE</td>
<td>6</td>
<td>Risk factors (blood pressure, cigarette smoking, cholesterol)</td>
<td></td>
</tr>
<tr>
<td>Hawkins</td>
<td>USA</td>
<td>Enoxaparin versus heparin for prevention of deep vein thrombosis after total hip replacement</td>
<td>CE</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kampulainen</td>
<td>Finland</td>
<td>Influenza vaccination of healthy employees</td>
<td>CE</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liem</td>
<td>Netherlands</td>
<td>Extraperitoneal laparoscopic inguinal hernia repair</td>
<td>CE</td>
<td>6</td>
<td>Age, gender, length, weight, number of patients with paid work</td>
<td></td>
</tr>
<tr>
<td>Marrazzo</td>
<td>USA</td>
<td>Selective screening criteria for Chlamydia trachomatis infection in women</td>
<td>CE</td>
<td>6</td>
<td>Age, ethnicity (white, black, Hispanic, Native American, Asian, other), behavioural characteristics</td>
<td></td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>First author</th>
<th>Country</th>
<th>Type of intervention</th>
<th>Study type</th>
<th>Outcome measure(^1)</th>
<th>Effects in different subpopulations</th>
<th>Comparison of CER by subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myers</td>
<td>USA</td>
<td>Fetal lung maturity testing in preterm labour</td>
<td>CE</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salign</td>
<td>France</td>
<td>Screening for HIV, hepatitis B, hepatitis C and human T-cell leukaemia virus in blood transfusions</td>
<td>CE</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taylor</td>
<td>USA</td>
<td>Treatments for the eradication of <em>Helicobacter pylori</em></td>
<td>CE</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visentin</td>
<td>Italy</td>
<td>Osteoporosis treatment</td>
<td>CE</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brannon</td>
<td>USA</td>
<td>Alternative methods of nutrition education for hypercholesterolaemic children</td>
<td>CE</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dranitsaris</td>
<td>Canada</td>
<td>Cancer support care</td>
<td>CB (WTP)</td>
<td>7</td>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>Goldstein</td>
<td>Canada</td>
<td>Respiratory rehabilitation</td>
<td>CE</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martin</td>
<td>Spain</td>
<td>Tuberculosis screening</td>
<td>CB (HC)</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O’Brien</td>
<td>Canada</td>
<td><em>Helicobacter pylori</em> eradication strategies in the management of duodenal ulcer</td>
<td>CE</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rissanen</td>
<td>Finland</td>
<td>Hip and knee replacement</td>
<td>CE</td>
<td>7</td>
<td>Age, gender, marital status, employment status</td>
<td>Age (&lt;60, 61–70, &lt;70 years)</td>
</tr>
<tr>
<td>Sadowski</td>
<td>Canada</td>
<td>Gastric ulcers at endoscopy</td>
<td>CE</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spearman</td>
<td>USA</td>
<td>Initial therapy with 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors to treat hypercholesterolaemia in the primary-care setting of a managed-care organisation</td>
<td>CE</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tang</td>
<td>China</td>
<td>Malaria surveillance in the late consolidation phase</td>
<td>CE</td>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^1\) CB, cost–benefit; CE, cost-effectiveness; HC, human-capital approach; WTP, willingness-to-pay approach

\(^2\) 1, Quality-adjusted life-expectancy; 2, survival rates; 3, cases treated successfully; 4, life-expectancy; 5, cases detected; 6, cases prevented; 7, other
Appendix 6

Cervical screening case study: sources of data and sensitivity analysis

Sources of effectiveness data

Age-specific annual incidence and mortality data were provided by the UK Office for National Statistics (partly published in Quinn and co-workers\textsuperscript{133}). Incidence data by 5-year age group for the year 1997 were estimated by applying the 1995 incidence/mortality ratio to 1997 mortality data. This method is similar to that adopted by the International Agency for Research on Cancer and by the UK Office for National Statistics, although we had to use a ratio based on a single year, rather than an average of 3 years, because the relevant information was only partially available. Age-specific incidence rates formed a basis for calculating the risk of developing invasive cervical cancer for women with intervals of varying length from their last smears. Relative risks were determined on the basis of the results of a case–control study carried out in England.\textsuperscript{159} Preference was given to this English data even if evidence from the USA\textsuperscript{133} and The Netherlands\textsuperscript{152} seems to suggest a higher risk reduction for screened women.

As previously mentioned, it has been argued that an underlying declining trend may exist in the incidence of invasive cervical cancer.\textsuperscript{130} The model adopted in this study shows the cost-effectiveness of alternative strategies at the current incidence rates (and risks of developing invasive cancer by time elapsed from the last smear, depending on coverage). If these results had to be projected into the future, alternative incidence rates should be tested. This was done as part of the sensitivity analysis, and results are discussed in the relevant section.

It is extremely difficult to estimate mortality rates and survival for women with cervical cancer at the current level of coverage. A declining trend in mortality over time has been shown by several studies.\textsuperscript{150,152,217} This has surely been affected by progressive increases in screening coverage, but survival rates of women diagnosed with invasive cancer seem to have improved as well,\textsuperscript{137} although this could also be partly due to screening, as cases tend to be diagnosed at an earlier stage. For the purposes of this study, gains in life-expectancy have been estimated by making projections of 5-year survival rates by age group for the cohort of women diagnosed with cervical cancer in the period 1990–95. Such projections have been based on the application of overall 5-year survival trends to age-specific figures for the cohort of women diagnosed with cervical cancer in the period 1986–90 (both available from Coleman and co-workers\textsuperscript{138}). It was assumed that women not surviving at 5 years have an average life-expectancy of 2 years, based on the available evidence about the distribution of cervical cancer deaths.\textsuperscript{157,218} The loss of life-expectancy for women in each 5-year age group was estimated from 1997 life-tables for England, and discounted at a 5% rate.

Sources of cost data

Most of the costs involved in running the screening programme were estimated on the basis of a recent National Audit Office report.\textsuperscript{135} This includes an assessment of invitation costs, GP target payments, and laboratory and colposcopy costs for the diagnosis and treatment of preinvasive forms of cervical cancer. Occasionally, the treatment of preinvasive cancer may require hospitalisation. Costs related to these rare occurrences have been neglected, consistently with the results of a cost-analysis of 141 cases, none of which required hospitalisation.\textsuperscript{219} Smear costs have been shown to vary widely between laboratories (£1.68 to £23.70), and therefore an average full cost of £8.86 per test was assumed, as reported by the National Audit Office.\textsuperscript{135}

The cost of treating invasive cancer was estimated on the basis of a study undertaken in the Trent region during the period 1990–95 on 261 women who developed invasive cervical cancer.\textsuperscript{219} The study provided stage-specific treatment costs, and the estimates incorporated in our model are averages of these costs weighted by the proportions of cases in each stage. This might lead to a slight overestimation of treatment costs, as the number of cases in the more advanced stages is likely to decrease when screening coverage is expanded (Wolstenholme
and Whynes's study refers to cancers developed in 1990, when coverage was 59%.

No information was available on the cost of increasing coverage beyond the current level. One possible strategy involving the use of health visitor follow-up of non-respondents was described, but no cost analysis was reported. Alternative strategies may be used for different target groups (e.g. low-income women, ethnic-minority women). A baseline extra-cost of £10 per additional woman screened was assumed, and the effect of possible higher costs, as suggested for example in a recent American study, was examined in the sensitivity analysis.

**Sensitivity analysis**

A number of one-way sensitivity analyses were carried out in order to assess the robustness of the results of the study to alternative assumptions regarding key variables. Most analyses determined only marginal changes in cost-effectiveness ratios and did not modify the ranking of alternative strategies, because changes in the variables tested affected both comparators.

The effects of the hypothesis of an underlying declining trend in the incidence of cervical cancer have been assessed by assuming a 10% decrease in the number of invasive cancers and deaths. This would lead to a 13% increase in the cost-effectiveness ratio of the current strategy compared to the previous strategy. The resulting ratio would still be largely acceptable. On the other hand, the cost-effectiveness of strategies aimed at improving the effectiveness of the current screening programme would become slightly better.

A further sensitivity analysis was conducted by replacing the cost figures derived from the National Audit Office report with unit cost data calculated for cervical screening in the Oxfordshire Region in 1992. Significant changes occurred, because National Audit Office did not include practice-nurse time and other primary-care costs in their analysis. If these costs were taken into account, the programme costs for all options would increase (by approximately £60 million for the current strategy). However, these costs are borne only in part by the NHS, and it is not known how the resources currently spent by general practices on screening-related activities would be used otherwise. Therefore, the actual economic impact of this additional factor is probably much more limited than the above study seems to suggest. Our estimates are consistent with those reported by Waugh and co-workers. As previously mentioned, the cost of reaching women currently not covered by the programme may vary significantly.

Finally, alternative scenarios were considered with regard to the screening option based on the situation prior to the introduction of target payments. In the baseline analysis a coverage of 59% was assumed for this hypothesis, corresponding to the level achieved at 31 March 1990. It may be argued that coverage rates might have increased further even without economic incentives to GPs. In fact, evidence from Italy and The Netherlands seems to indicate that spontaneous screening would be extremely common and coverage would be high in the absence an organised programme, the estimated contribution of the latter being 17% and 24%, respectively. Therefore, a simulation was carried out aimed at determining the incremental cost-effectiveness of the current strategy compared to hypothetical situations in which a coverage of 70% and 75% (all women screened at 5-year intervals), instead of 59%, were reached without target payments. The resulting cost-effectiveness ratios were higher than in the baseline analysis, but still within an acceptable range (approximately £18,400 and £22,800, respectively).
Methodology Group

Members

Methodology Programme
Director
Professor Richard Lilford
Director of Research
and Development
NHS Executive – West
Midlands, Birmingham

Chair
Professor Martin Buxton
Director, Health Economics
Research Group
Brunel University, Uxbridge

Professor Ann Bowling
Professor of
Health Services Research
University College London
Medical School

Professor David Chadwick
Professor of Neurology
The Walton Centre for
Neurology & Neurosurgery
Liverpool

Dr Mike Clarke
Associate Director (Research)
UK Cochrane Centre, Oxford

Professor Paul Dieppe
Director, MRC Health Services
Research Centre
University of Bristol

Professor Michael Drummond
Director, Centre for
Health Economics
University of York

Dr Vikki Entwistle
Senior Research Fellow,
Health Services Research Unit
University of Aberdeen

Professor Ewan B Ferlie
Professor of Public Services
Management
Imperial College, London

Professor Ray Fitzpatrick
Professor of Public Health
& Primary Care
University of Oxford

Dr Naomi Fulop
Deputy Director,
Service Delivery &
Organisation Programme
London School of Hygiene
& Tropical Medicine

Mrs Jenny Griffin
Head, Policy Research
Programme
Department of Health
London

Professor Jeremy Grimshaw
Programme Director
Health Services Research Unit
University of Aberdeen

Professor Stephen Harrison
Professor of Social Policy
University of Manchester

Mr John Henderson
Economic Advisor
Department of Health, London

Professor Theresa Marteau
Director, Psychology &
Genetics Research Group
Guy’s, King’s & St Thomas’s
School of Medicine, London

Dr Henry McQuay
Clinical Reader in Pain Relief
University of Oxford

Dr Nick Payne
Consultant Senior Lecturer
in Public Health Medicine
ScHARR
University of Sheffield

Professor Joy Townsend
Director, Centre for Research
in Primary & Community Care
University of Hertfordshire

Professor Kent Woods
Director, NHS HTA
Programme, & Professor
of Therapeutics
University of Leicester

Professor Ray Fitzpatrick
Professor of Public Health
& Primary Care
University of Oxford

Professor Theresa Marteau
Director, Psychology &
Genetics Research Group
Guy’s, King’s & St Thomas’s
School of Medicine, London

Dr Henry McQuay
Clinical Reader in Pain Relief
University of Oxford

Dr Nick Payne
Consultant Senior Lecturer
in Public Health Medicine
ScHARR
University of Sheffield

Professor Joy Townsend
Director, Centre for Research
in Primary & Community Care
University of Hertfordshire

Professor Kent Woods
Director, NHS HTA
Programme, & Professor
of Therapeutics
University of Leicester

Current and past membership details of all HTA ‘committees’ are available from the HTA website (see inside front cover for details)
Current and past membership details of all HTA ‘committees’ are available from the HTA website (see inside front cover for details)
Feedback

The HTA programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (http://www.ncchta.org) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

We look forward to hearing from you.