

# **Making the most of it**

**Economic evaluation in the social welfare field**

**Tom Sefton, Sarah Byford, David McDaid, John Hills and  
Martin Knapp**

The **Joseph Rowntree Foundation** has supported this project as part of its programme of research and innovative development projects, which it hopes will be of value to policy makers, practitioners and service users. The facts presented and views expressed in this report are, however, those of the authors and not necessarily those of the Foundation.

Joseph Rowntree Foundation, The Homestead, 40 Water End, York YO30 6WP  
Website: [www.jrf.org.uk](http://www.jrf.org.uk)

This project was carried out jointly by the London School of Economics (LSE) and the Institute of Psychiatry (IOP). **Tom Sefton** is a Research Officer in the ESRC Research Centre for Analysis of Social Exclusion (CASE) at the LSE, **Sarah Byford** is Senior Lecturer in the Centre for the Economics of Mental Health at the IOP, **David McDaid** is a Research Officer in LSE Health and Social Care, **John Hills** is Professor of Social Policy at LSE and Director of CASE, and **Martin Knapp** is Professor of Social Policy at LSE and Director of the LSE branch of the Personal Social Services Research Unit.

© Joseph Rowntree Foundation 2002

First published 2002 by the Joseph Rowntree Foundation

All rights reserved. Reproduction of this report by photocopying or electronic means for non-commercial purposes is permitted. Otherwise, no part of this report may be reproduced, adapted, stored in a retrieval system or transmitted by any means, electronic, mechanical, photocopying, or otherwise without the prior written permission of the Joseph Rowntree Foundation.

ISBN 1 84263 070 9 (paperback)  
ISBN 1 84263 132 2 (pdf: available at [www.jrf.org.uk](http://www.jrf.org.uk))

A CIP catalogue record for this report is available from the British Library.

Prepared and printed by:  
York Publishing Services Ltd  
64 Hallfield Road  
Layerthorpe  
York YO31 7ZQ  
Tel: 01904 430033; Fax: 01904 430868; Website: [www.yps-publishing.co.uk](http://www.yps-publishing.co.uk)

*Further copies of this report, or any other JRF publication, can be obtained either from the JRF website ([www.jrf.org.uk/bookshop/](http://www.jrf.org.uk/bookshop/)) or from our distributor, York Publishing Services Ltd, at the above address.*

# Contents

<b>Acknowledgements</b>	v
<b>PART I: BACKGROUND</b>	1
<b>1 Introduction</b>	3
<b>2 What is economic evaluation?</b>	7
Purpose	7
Definition	7
Principles of economic evaluation	8
Types of economic evaluation	9
Building blocks	11
<b>3 Current state of economic evaluation</b>	15
Methodology of systematic review	15
Results	15
<b>PART II: CONCEPTUAL ISSUES</b>	19
<b>4 Challenges for economic evaluation</b>	21
Why so little economic evaluation?	21
Nature of programmes	21
<b>5 Alternative perspectives on evaluation</b>	27
Different schools of thought	27
Strengths and limitations of conventional approaches	29
Possible framework for evaluating complex interventions	30
<b>PART III: PRACTICAL APPLICATION</b>	33
<b>6 Clarifying purpose and objectives</b>	35
Key aspects of evaluation	35
Being clear about the purpose of evaluation	35
Clarifying programme objectives	38
<b>7 Assessing outcomes and costs</b>	41
Choosing the right outcomes	41
Measuring outcomes	45
Assessment of costs	51
Conclusions	54
<b>8 Study design and methods</b>	56
Study design	56
Complementary research methods	70
Other design issues	75

<b>9</b>	<b>Analysing and interpreting results</b>	81
	Analysis of results	81
	Understanding how and why programmes work	83
	Interpretation and presentation of results	88
<b>10</b>	<b>Managing the evaluation process</b>	90
	Involving key stakeholders	90
	Working with other disciplines	91
	Constraints on evaluation	93
<b>11</b>	<b>Summary and conclusions</b>	96
	<b>Note</b>	99
	<b>References</b>	100

# Acknowledgements

This project was carried out by the ESRC Research Centre for Analysis of Social Exclusion (CASE) and the Personal Social Services Research Unit (PSSRU) at the London School of Economics and by the Centre for the Economics of Mental Health at the Institute of Psychiatry. It was initiated and funded by the Joseph Rowntree Foundation. The research team is grateful for this support and for the help and advice of Janet Lewis, who also chaired the Advisory Group meetings.

The authors would like to thank the rest of the research team for their significant contributions to the final report. Jennifer Beecham, Howard Glennerster, Ann Netten and Franco Sassi contributed their expertise on different aspects of evaluation and offered many useful and detailed comments on the final report. Other researchers in CASE were closely involved in some of the case study evaluations, including Simon Burgess, Liz Richardson and Becky Tunstall.

The authors would also like to thank members of the external Advisory Group: Martin Buxton, James Foreman-Peck, Sally Holtermann, Robert Maxwell, Tim Newburn, David Pearce, Becky Sandhu and Elliot Stern. These meetings generated a stimulating debate on economic evaluation and highlighted the differences in perspective between economists and non-economists and also within the economics field. It is hoped this report has gone some way to exposing and addressing, though not always resolving, many of these issues. The authors would particularly like to thank Elliot Stern who encouraged us early on to look beyond the economics literature on evaluation and Sally Holtermann who provided detailed comments on the final report and whose earlier report, *Weighing it up*, was one of the catalysts for this project.



# **Part I**

## **Background**





# 1 Introduction

## Purpose

This report is the main output of a two-year project to promote better understanding and use of economic evaluation in the social welfare field. The need for economic evaluation is not always acknowledged. Where it is recognised, there is often a large gap between the expectations of those commissioning evaluation research and what can realistically be delivered as part of an economic evaluation. This report does not seek to be prescriptive about the methods used, but to highlight some of the key issues for those evaluating social welfare programmes and how these might be addressed.

The report is aimed at both economists and non-economists involved in the evaluation of social welfare programmes, but who are frustrated by some of the problems of carrying out good economic evaluation in practice. It assumes a basic understanding of economic evaluation, so key concepts are defined only briefly, but we have kept the use of economic jargon to a minimum. The focus of this report is on the application of these approaches in a social welfare context. The accompanying report (Volume II) provides more detailed technical advice on how some of these principles might be applied in different situations (Byford *et al.*, 2002, forthcoming). Many of the issues discussed in this report are not specific to economic evaluation and are relevant to those with an interest in evaluation more generally.

## Scope

Economic evaluation is defined here as the systematic attempt to identify, measure and compare the costs and outcomes of alternative

interventions. Our focus is on economic evaluation at the practice level, as opposed to broader or more strategic policy questions.

Our definition of social welfare encompasses social care, early intervention schemes, housing, regeneration, community development, work with children, young people and families and welfare to work and related areas. This broadly corresponds with the areas covered by the Joseph Rowntree Foundation's research priorities (Joseph Rowntree Foundation, 2001a). We have excluded health care because economic evaluation is already well developed in this area. Nor did we consider mainstream education, environment, or transport, although many of the issues will be similar. Within the areas we have looked at, we have placed greater emphasis on more complex interventions, where the biggest gaps are and where we felt our input would be more valuable.

## Our perspective

We started by analysing why there is relatively little economic evaluation in the social welfare field. We looked at the reasons why the evaluation of these programmes can be so challenging. We also examined the methods employed in the health care field, where the use of economic evaluation is much more widespread, to see whether similar approaches could be transferred to other areas of social welfare. This starting point explains our emphasis on experimental approaches to evaluation, including randomised controlled trials and other more pragmatic experimental designs. We believe these are generally more applicable to the social welfare field than the econometric or modelling techniques that are

dominant in certain other policy areas, such as transport. However, it is important to consider how experimental approaches may need adapting in a social welfare context, because these interventions are often very different to the kinds of interventions that most health economists have focused on.

We do not believe that the only worthwhile studies are quantitative analyses carried out on the back of large-scale randomised controlled trials or that economists have little to offer in the absence of this type of analysis. But, we also do not believe that rigorous quantitative approaches to evaluation are generally infeasible or inappropriate in the social welfare field. We argue that economists should be attempting to carry out systematic assessments of costs and outcomes and that study design is important, though not exclusively so. Other issues, such as clarifying the objectives of a programme and choosing the right outcomes, are also central, crucial, and sometimes neglected, aspects of economic evaluation.

We are pragmatic in our approach to economic evaluation, recognising the practical and methodological problems of evaluating complex social welfare interventions. For this reason, we believe that the use of complementary research methods would improve the validity of studies and help to interpret their results. The report is eclectic, rather than dogmatic in its approach, drawing from a wide range of perspectives on evaluation, including the non-economics literature.

It became clear early on that this was a two-way learning process. Economists have a valuable contribution to make to evaluation in the social welfare field, but they also have plenty to learn from other perspectives on

evaluation. We believe strongly that a multi-disciplinary approach is the best way forward, though there is a long way to go in achieving this ideal.

### Approach

We examined the evaluation literature to identify the main conceptual and methodological issues for economic evaluation and for evaluation more broadly. We focused specifically on the implications for evaluating social welfare interventions, especially more complex programmes.

However, we did not want this report to be a purely theoretical discussion of the issues, so we carried out several case study evaluations. These forced us to confront many of the obstacles to economic evaluation in this field. The case studies, which are summarised in the next chapter, span a wide range of social welfare areas: community development, youth homelessness prevention, foster care and fuel poverty. They are used to illustrate many of the points we make; they also gave us an insight into the process of conducting economic evaluation, which we could not have gained by reading the reports of previous studies.

In addition, we carried out an electronic search of studies published in academic journals to provide evidence on the current state of economic evaluation in the social welfare field and to identify examples of good practice across a broad spectrum of interventions.

### Structure of the report

Chapter 2 provides a definition of economic evaluation and outlines the main types. It

recognises the difficulty of conducting a full economic evaluation in many areas of social welfare and identifies the building blocks that are needed to carry out rigorous economic evaluation. This chapter is directed primarily at non-economists and can be skipped by those familiar with the principles of economic evaluation. Chapter 3 presents the results of our systematic review of economic evaluation studies.

Part II (Chapters 4 and 5) explores certain conceptual issues and their implications for economic evaluation. Chapter 4 considers the challenges faced by those attempting to evaluate complex social welfare interventions. Chapter 5 introduces some alternative perspectives on evaluation, which have shaped our approach to economic evaluation in this report.

Part III of the report (Chapters 6 to 10) discusses the practical application of economic evaluation in the social welfare field with examples from our case studies and other published studies. These chapters are in a logical order, but can also be used as a reference guide for those interested in specific aspects of evaluation.

After a brief introduction to the key aspects of economic evaluation, Chapter 6 considers the different purposes of evaluation and some of the implications for economic evaluation. It stresses the importance of clarifying a programme's objectives as a fundamental and substantive part of any evaluation.

Chapter 7 provides guidance on identifying and assessing relevant costs and outcomes. It emphasises the need to choose outcomes carefully and looks at ways of measuring them, including those that are not naturally amenable to measurement. It also argues that the costs

side of economic evaluation needs to be given more attention than is often the case, taking a broader perspective on costs and treating cost analysis as a more integral part of evaluation.

Chapter 8 discusses the main approaches to economic evaluation, focusing on experimental designs with random and non-random comparison groups. The more complex the intervention, the more likely it is that experimental approaches will be insufficient on their own. Other research methods are, therefore, discussed, including the use of qualitative methods, which we believe have an important role to play in economic evaluation. Other design issues, such as sample size and follow-up, are considered, too.

Chapter 9 discusses the final stages of economic evaluation, where data on costs and outcomes are analysed and, where appropriate, combined to make a judgement about the cost-effectiveness of an intervention. A strong theme running through this report, which is examined in more detail in this chapter, is that economists should place more emphasis on understanding how and why programmes work (or not). The final section of this chapter provides some general guidelines for presenting the results of economic evaluations.

Chapter 10 discusses various issues concerned with the evaluation process, which are sometimes neglected by evaluators who may be preoccupied with evaluation methodology. Our case studies have been particularly helpful in highlighting the importance of involving other stakeholders at different stages in the evaluation, the need to work with evaluators from other disciplines, and, not least, the financial and other constraints on evaluation. This comes at the end

## Making the most of it

---

of the report, because it refers back to the discussion in previous chapters; however, these issues need to be addressed early on in an evaluation study.

Finally, Chapter 11 summarises the arguments in the report on developing the use of economic evaluation in the social welfare field.

## 2 What is economic evaluation?

### Purpose

The purpose of economic evaluation is to inform decisions about the best use of limited resources. Using resources for a particular intervention means fewer resources will be available for other uses. A service may be effective to a degree in meeting the needs of its users, but it is not cost-effective if similar outcomes could be achieved at less cost, or if better outcomes could be achieved for the same cost. Most evaluation focuses on outcomes. Economic evaluation takes into account both the costs and outcomes of an intervention with a view to identifying the most efficient way of achieving policy objectives.

While all policy decisions are heavily influenced by political, ethical and pragmatic considerations, economic evaluation has a potentially valuable role to play in this process. Decisions of this kind have to be made at all levels: whether it is about the allocation of spending between programme areas, whether to implement a new scheme, or what level of service to provide to individual users. The purpose of economic evaluation is to enable these decisions to be made on a more rational basis. This is consistent with the increased emphasis being placed on evidence-based policy-making both within and outside government (see, for example, Centre for Management and Policy Studies, 2001).

### Definition

The most common definition of economic evaluation, and the one we use, is a systematic attempt to identify, measure and compare the costs and outcomes of alternative interventions (Drummond *et al.*, 1997; HM Treasury, 1997).

Breaking this down, the basic requirements for economic evaluation are:

- focus on a specific intervention
- systematic assessment of costs and outcomes
- some sort of benchmark or point of comparison
- attempt to combine costs and outcomes in the final analysis.

Economists have an important contribution to make at all stages of the evaluation process, including:

- helping to clarify objectives and convert these into outcomes that are measurable
- drawing a clear distinction between processes, inputs, outputs and outcomes
- encouraging a more systematic and rigorous assessment of costs and outcomes, with a particular emphasis on generating statistically valid results
- highlighting the need to consider what would have happened in the absence of the intervention being evaluated
- adopting a societal perspective or multiple perspectives, thus ensuring a more comprehensive assessment of a programme's impact
- helping to explain how policies or programmes operate, especially those that are economic by nature or have an economic element to them (e.g. a tax incentive to encourage people to move into employment).

Thus, we do not believe that the assessment of outcomes is outside the domain of economics or that the role of economists should be confined to the valuation of outcomes. We would argue that economists ought to be involved in helping to assess outcomes and that good economic evaluation can be carried out without necessarily seeking to value outcomes in monetary terms.

### Exclusions

Economic evaluation is not about the financial viability of a programme or the organisation responsible for delivering it. It is possible to have an organisation that is struggling to make ends meet, but that is delivering a cost-effective programme, once the benefits to users and to society at large are taken into account.

Another common misconception is that economic evaluation is only about costs, which is why the role of economists is often restricted to carrying out stand-alone cost studies. Information on costs is a vital part of economic evaluation. Cost studies can also be useful in their own right (Holtermann, 1998), for example as part of a budgeting exercise or to inform a broad policy debate (see, for example Holtermann, 1995). However, they do not tell you whether a particular programme is worthwhile; for this, it is necessary to look at both costs and outcomes together.

We also attempt to make a distinction between economic evaluation and other types of economic analysis, though this can be somewhat blurred in practice. Microeconomics is concerned with the economic behaviour of people and organisations: how they make choices to consume or provide certain goods or services, and the outcomes of all these

individual decisions. For example, economists have developed theories and models to help understand individuals' job-seeking behaviour. While this kind of economic analysis has implications for the cost-effectiveness of certain types of intervention and could, therefore, feed into an evaluation (e.g. of a welfare-to-work scheme), we would not include this in our definition of economic evaluation, because it does not focus on evaluating a specific intervention.

### Principles of economic evaluation

Economic evaluation is underpinned by a number of fundamental principles derived from welfare economics:

- *Comparative*: since economics is concerned with the best use of scarce resources, economic analysis must involve comparisons between alternative courses of action.
- *Societal perspective*: decisions about resource allocation should be based on an assessment of all the possible costs and benefits of a programme to all sectors of society (Johannesson, 1995) and their distributional effects. (There may, however, be circumstances where a narrower perspective is appropriate, for example to understand the incentives facing a particular agency.)
- *Final outcomes*: the aim of welfare economics is to maximise the well-being of society. For this reason, economists prefer to focus on final outcomes (i.e. changes in quality of life), rather than

intermediate outcomes (e.g. better housing services) or process measures (e.g. more tenant consultation).

- *Preference-based*: cost-benefit analysis (and welfare economics, more generally) identifies well-being with the satisfaction of consumer preferences (Hausman and McPherson, 1994). Hence, economists place great emphasis on individuals'

preferences in assessing the value of different outcomes.

### Types of economic evaluation

There are different types of economic evaluation. While these all involve a systematic attempt to assess relevant costs and outcomes, they differ in their approach to measuring outcomes. The main types are described in Box 1.

#### Box 1 Methods of economic evaluation

##### *Cost-minimisation analysis*

This is the simplest form of economic evaluation, but should be used only when there is strong evidence that the interventions under consideration are equally effective. This type of evaluation involves the comparison of costs alone in order to identify the least cost alternative.

##### *Cost-effectiveness analysis*

Cost-effectiveness analysis (CEA) is the most common approach to economic evaluation in health care and involves the measurement of benefits in 'natural' units such as life years gained, number of disability-free days, or levels of depression. In social welfare research, natural units would include outcomes such as improvements in family functioning or young people's self-esteem. The benefits of two or more interventions are combined with their respective costs, providing a measure of cost per unit. Priority should be given to those interventions with the lowest cost per unit of outcome gained.

##### *Cost-utility analysis*

This is a specific form of cost-effectiveness analysis (CEA). As with CEA, it provides a measure per unit of improvement in outcome, but outcomes are measured in terms of utility (or well-being). One example of a utility measure is the Quality Adjusted Life Year (or QALY). This involves the use of weights to adjust for differences in the perceived quality of life associated with different health states. These weights are usually based on preference surveys of the general population.

##### *Cost-consequences analysis*

This is similar to cost-effectiveness analysis, except that there is more than one outcome measure. For some interventions, it would be inappropriate to focus on a single outcome, because this would omit other important outcomes from the analysis. These studies present the outcomes of an intervention without attempting to combine them into a single measure of

*(continued overleaf)*

effectiveness. Cost-consequences analysis does not allow different interventions to be ranked in terms of their efficiency, unless one intervention is no more expensive and performs better on all the outcomes measured. But, it does allow decision-makers to make a more informed judgement, based on their own implicit or explicit weighting of different outcomes.

### *Cost-benefit analysis*

Cost-benefit analysis (CBA) requires both costs and benefits to be measured in monetary amounts, so it is possible to make a direct comparison between the costs and benefits of an intervention. A project or programme is worthwhile if the benefits exceed the costs, subject to any budget constraint. This approach allows comparisons to be made between interventions across different sectors, such as health, education, or defence. The problem, however, is how to value the benefits in monetary terms. Methods do exist, but these are controversial and not widely used except in a few specific areas, such as transport.

Some studies evaluate the effectiveness of a programme in reducing a social 'bad' (e.g. crime), and then estimate the potential savings associated with this outcome (e.g. reductions in expenditure on the criminal justice system). Although these are often presented as cost-benefit analyses, they are more accurately described as *cost-savings or cost-offset analyses*, because this approach to valuing benefits is not consistent with the principles of welfare economics. For example, the amount spent on prosecuting and incarcerating offenders is not the same as the amount people would be willing to pay to cut crime (which is what economists would be seeking to value in a CBA). Furthermore, these studies are rarely comprehensive, because they tend to ignore benefits that do not generate tangible savings (e.g. psychological costs to victims of crime). Nevertheless, cost-savings analyses can be useful in making the case for a particular programme if the expected savings can be shown to be greater than expected costs.

The feasibility of each type of evaluation depends on the nature of outcomes and the scope for combining or monetising benefits. Cost-consequences analysis is currently the dominant type in the social welfare field (see Chapter 3). Cost-effectiveness analyses are less common, because it is not usually appropriate to summarise the impact of social welfare programmes using a single outcome measure. Cost-benefit analysis is difficult, because many outcomes, especially in social welfare, are hard to value in monetary terms (see section on 'Measuring outcomes' in Chapter 7).

While feasibility may determine the choice of approach, this should also depend on the purpose of the study. For example, if the aim of an evaluation is to compare two programmes that have similar objectives, then there may be little to be gained from monetising the benefits; hence a cost-effectiveness or cost-consequences analysis is likely to be more appropriate. A good example would be the evaluation of a group-based early reading programme that examines the impact on test scores compared with more expensive one-to-one reading programmes (Nicolson *et al.*, 1999). The results showed that



the improvement in reading scores was comparable to those of the established Reading Recovery Programme, but at around 10 per cent of the cost. Thus, a cost-effectiveness study was sufficient to demonstrate the (relative) efficiency of this programme, compared with another established programme.

If the aim were to assess whether a new programme represented a good use of public money compared with programmes in other areas, then cost-benefit analysis would be preferable, provided that all the important benefits could be valued in monetary terms. A good example is an evaluation of a ten-year programme of energy efficiency measures in Ireland (Clinch and Healy, 2001). This study consists of a broad-brush, but comprehensive, cost-benefit analysis of the programme, covering potential energy savings, environmental benefits and improvements in comfort and health, all of which are valued in monetary terms. The results show that there are clear net benefits from such a programme. Cost-benefit analysis is useful in making a general case for this kind of programme where there is no clear alternative against which comparisons can be made.

Another important distinction is between evaluation that is carried out *ex ante* (i.e. before a programme is implemented), prospectively (i.e. alongside the implementation of a programme), or retrospectively. *Ex ante* evaluation, which is often termed appraisal (HM Treasury, 1997), is a useful exercise that can help to inform a decision about whether to go ahead with a new programme and, if so, how to target a programme to make it more cost-effective. But, appraisal is not a substitute for evaluating a programme once it has been implemented.

Ideally, all evaluation should be carried out

prospectively, so that the programme can be monitored, and if necessary modified, from an early stage. This also helps to ensure that the necessary data on costs and outcomes are collected. Retrospective evaluation is clearly better than no evaluation, but these studies are often constrained by poor data. Often, this information has been collected for other (e.g. administrative) reasons and may be inappropriate or incomplete for evaluation purposes. Certain study designs, including randomised controlled trials, can only be carried out prospectively.

### Building blocks

It is important to recognise the difficulties of carrying out a full and rigorous economic evaluation in many areas of social welfare. The development of economic evaluation in these areas will be a gradual process that gains momentum as new studies build on previous work. It is helpful to think in terms of the building blocks that are required to enable a fuller economic evaluation to be carried out at a later stage (see Figure 1). These preparatory stages are also useful to decision-makers in their own right.

We conducted four case studies as part of this research project, focusing on different aspects of economic evaluation. Although none of these represents a complete economic evaluation, they seek to put in place some of the building blocks of economic evaluation. Box 2 provides a brief description of each case study. More detailed information on the case studies is available from the project website ([www.econeval-social-welfare.org](http://www.econeval-social-welfare.org)). Many of the evaluation issues they raise are followed up throughout the report in the various boxes that accompany the main text.

### Box 2 Description of case study evaluations

#### *Safe In The City case study*

The Safe In The City (SITC) programme funds and monitors a mix of projects directed at young people who are seen to be at risk of homelessness. It has been operating for around three years. The second phase of projects, which began in 2001, is organised into eight geographical clusters. Each cluster has one project related to personal development, one related to employment and skills, and one that involves working with families. Once referred to the programme, clients have access to all of the projects within that cluster.

The purpose of our work was to help SITC to develop an evaluation strategy for their youth homelessness prevention programme. Our research consisted of:

- a selective review of the evaluation literature, focusing on programmes relating to homelessness, youth work and family services
- structured interviews with project managers and workers from nine projects funded by SITC
- discussions with SITC staff and with experts in this field.

This case study examines the feasibility of an experimental evaluation of the second phase of the programme and advises on the most appropriate methods to employ, including issues of study design, outcome measurement and assessment of costs. The nature of the SITC programme presents special challenges for this kind of evaluation. It was, therefore, a good opportunity to explore these issues, which have wider relevance for other social welfare programmes.

#### *Fuel poverty case study*

The Home Energy Efficiency Scheme (also called 'Warm Front') was launched in June 2000 and is designed to tackle fuel poverty among those most vulnerable to cold-related ill health. Low-income families and disabled people are eligible for grants of up to £1,000 for specified insulation measures. Low-income older people are eligible for grants of up to £2,000 for better insulation and improved heating systems. The scheme is targeted at private sector housing where nearly 70 per cent of fuel-poor households are found. The Government has allocated over £600 million to this scheme up to 2004, by which time the scheme is expected to have benefited around 800,000 vulnerable households.

The main purpose of this case study was to model the impact of the Home Energy Efficiency Scheme (HEES) within an economic framework, using data from a large-scale household survey (which contains relevant information on over 7,000 households). This was used to explore the cost-effectiveness of the scheme, in particular how it should be targeted to achieve the greatest reduction in the number of people living in fuel poverty. There were three principal components to this research:

- an econometric analysis, exploring the relationship between the warmth of people's homes and the energy efficiency of their home (the 'demand for warmth' equation)
- in-depth interviews with 18 households who have applied for HEES to test some of the assumptions underlying the 'demand for warmth' equation
- a simulation model to evaluate the impact of the HEES scheme, building on the results of the econometric analysis.

### *Gatsby case study*

This is a five-year programme funded by the Gatsby Charitable Foundation to stimulate small community self-help projects across the UK, mostly on deprived council estates. The Gatsby scheme offers short residential training courses for community groups. Participants develop their own ideas for community projects and can apply for grants of up to £5,000. Since 1996, around 700 groups have attended one of these training courses and over 70 grants have been awarded to a wide range of projects, including community gardens, after-school clubs, community cafes and advice centres. The distinctive aspects of the scheme are the strong emphasis on training, the link between training and the grants scheme, and the ongoing support provided to grant recipients.

This case study supports a piece of 'action research' being carried out at the London School of Economics (LSE), by seeking to develop a more systematic framework for assessing outcomes. Our work involved:

- selectively reviewing the literature on current approaches to evaluation in the field of community development
- exploring theories about what makes an effective community group, based on the literature and LSE's own experience of working with community groups
- developing a set of qualitative and quantitative methods for assessing community groups against these criteria
- testing out this approach on a small number of community groups.

By seeking to identify relevant outcomes and analyse these on a more consistent basis, our research provides one of the key building blocks needed for an economic evaluation of the scheme.

### *Foster care case study*

This study is a two-year follow-up of a large group of children who were originally in foster care. It aims to explore and explain their subsequent care careers, outcomes and costs with particular reference to the placement types that these young people have experienced since initial contact with the researchers. The results will not be able to provide a definitive answer to questions about the relative cost-effectiveness of alternative placements, but will offer evidence

*(continued overleaf)*

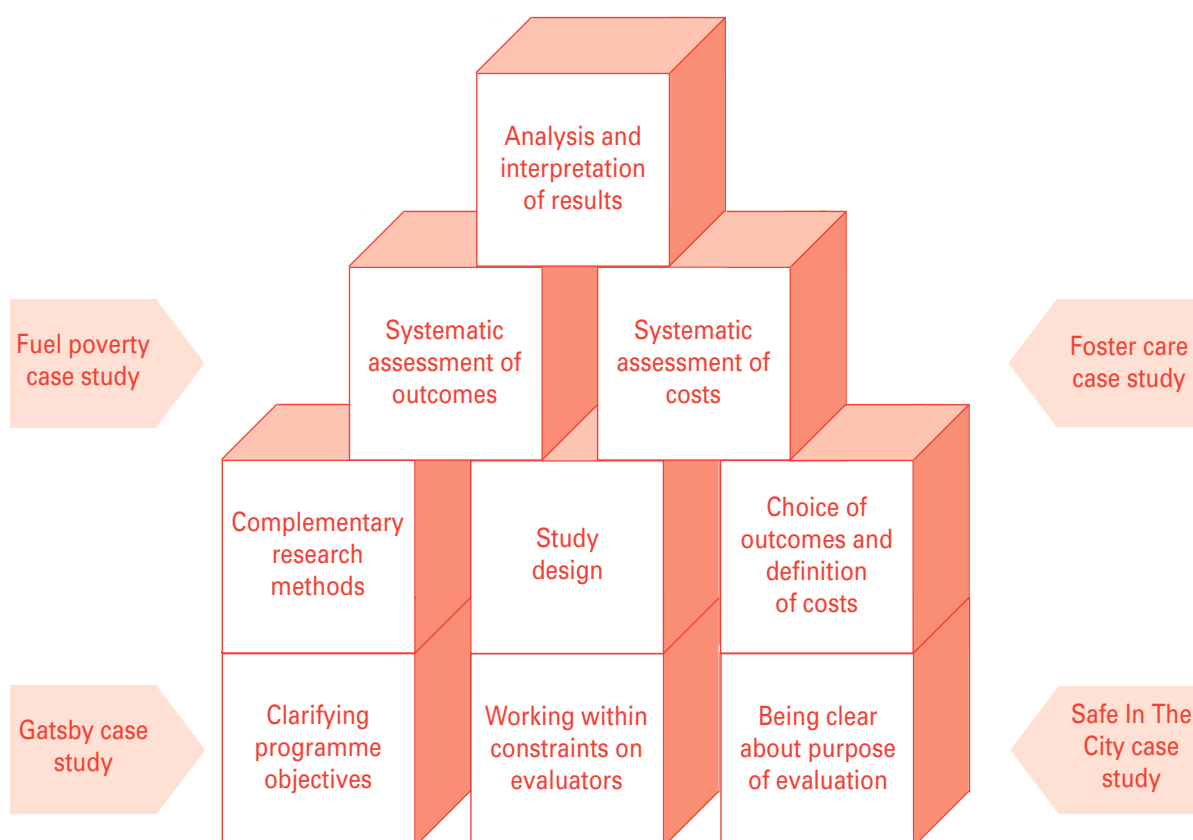
## Making the most of it

---

to help guide choices in the future and gain a deeper understanding of the characteristics and support services that predict good outcomes and indicate greater cost-effectiveness for different types of young people. The aims were to:

- describe the outcomes of placements (where the children are, how they are, and the fit between the placement and the wishes of the parties involved)
- explain the children's care careers and their continuing contact or lack of it with foster carers after they have left the care system
- evaluate the impact of placements on children's well-being with reference to placement type, support given by social services and the approach of carers
- estimate the costs of social services care packages and estimate the wider cost to other sectors in society in order to explore the cost-effectiveness of alternative placement types.

**Figure 1 Building blocks for economic evaluation**



### 3 Current state of economic evaluation

One of the motivations behind this project was that there appeared to be relatively little economic evaluation in many areas of social welfare, at least in the UK. We carried out a systematic review of economic evaluation studies in the social welfare field in order to provide more concrete evidence on this. We were also interested in the prevalence of studies in different areas of social welfare; we already knew, for example, that health care was an exception. This analysis is reported in more detail in a separate publication (McDaid *et al.*, 2002, forthcoming).

#### Methodology of systematic review

Our systematic review of published studies was based on a search of electronic databases, covering the major international journals for the period 1996–2000, but excluding unpublished studies or studies published outside journals, including government reports. The search strategy combined a set of economic evaluation terms (e.g. cost-effectiveness, cost-benefit) with a set of social welfare terms (e.g. social care, families, housing). The resulting list of abstracts was ‘filtered’ and assessed by the research team, using an agreed set of criteria to identify and classify economic evaluations. Information was recorded on the type of economic evaluation, the area of social welfare and the country of origin.

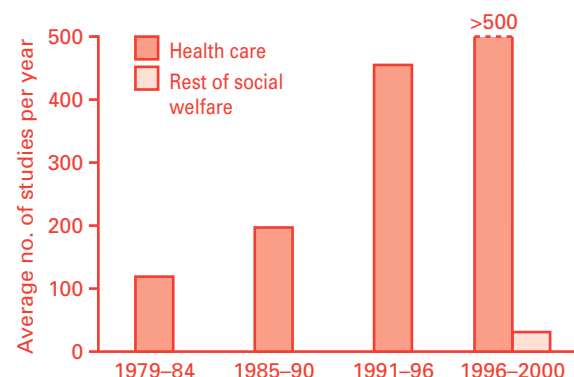
#### Results

Over the five-year period, we found a total of 131 economic evaluations. Allowing for the fact that electronic records were not complete for the most recent year, this is an average of around 30

studies per year. A similar review for the health care sector identified 450 studies per year over the period 1991–96 (Elixhauser *et al.*, 1998) and this is likely to have increased significantly since then. Figure 2 shows that there is a clear upward trend in the number of health care studies over time, based on an earlier review using the same approach (Elixhauser *et al.*, 1993). By contrast, there is no evidence that the number of studies is rising in the rest of the social welfare field, at least over the five-year period we examined.

Figure 3 shows the number of studies by area of social welfare. These appear to be concentrated in those areas most closely related to mainstream health, in particular mental health and public health (which includes drug and alcohol abuse). Together, these account for two-thirds of all the studies on our database. We also found that many studies (around a third) covered more than one of our categories. For example, there were supported employment schemes for people with mental health problems and drug treatment programmes that were also aimed at reducing crime or homelessness. This reflects the

Figure 2 Total number of economic evaluation studies



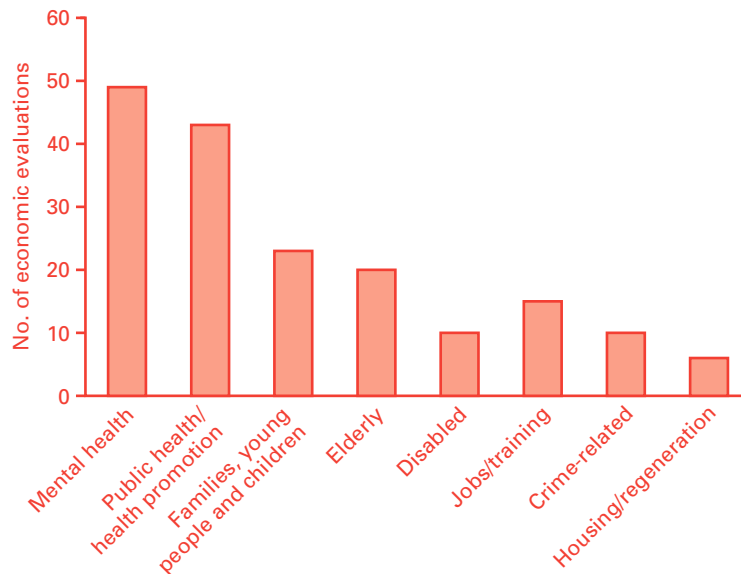
interrelationships between many areas of social welfare and contributes towards the complexity of these programmes.

As we expected, the majority of studies are cost-consequences analyses with multiple outcomes (see Figure 4). Less than one in 20 are cost-benefit analyses, excluding cost-savings analyses (see Box 1). Interestingly, the proportion of cost-benefit studies is also low in

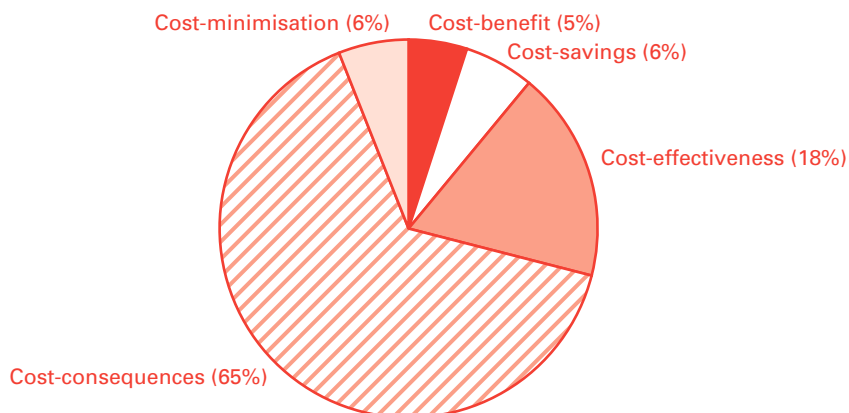
the health care sector and has been falling (from over 40 per cent in 1979–84 to under 20 per cent in 1991–96). It seems likely that a higher proportion of health care studies are cost-effectiveness analyses than in the social welfare field.

Nearly three-quarters of all the studies we identified are from the US (see Figure 5). The remainder are split fairly evenly between the

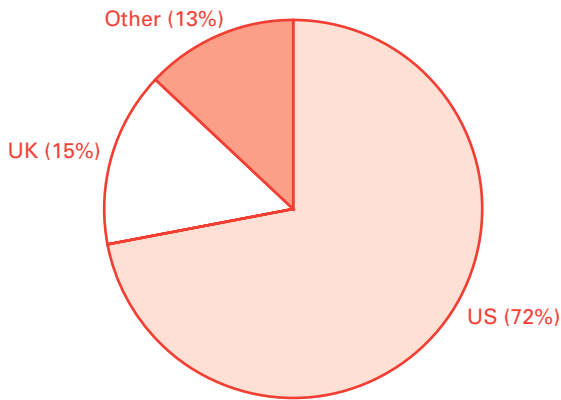
**Figure 3 Number of studies by area of social welfare (1996–2000)**



**Figure 4 Percentage of studies by type of study (1991–96)**



**Figure 5 Percentage of studies by country of origin (1991–96)**



UK and the rest of the world, although there may be some bias here because we were only able to consider abstracts in English, French or Spanish. The dominance of US studies is greater than in the health care sector, where just over half the studies were from the US (over the period 1991–96).

### Quality of studies

The quality of economic evaluations is very difficult to judge on the basis of the abstract alone, although in most cases we had information on sample size and the type of comparator, which gives us some idea of the scale and scientific rigour of these studies. Around a third of studies had no formal comparison group and one in six had a sample size of less than 50 (although this is not necessarily a problem). We also obtained the full paper for around 40 of the 131 studies and so were able to make a more accurate assessment of their strengths and weaknesses. There were examples of good practice, which we refer to in subsequent chapters. However, the quality of

these studies was very variable. The main weaknesses that we identified were as follows:

- Few studies included a comprehensive assessment of costs. Even when outcomes had been assessed rigorously, the analysis of costs was often crude.
- Yet many of the same studies made strong claims about the cost-effectiveness of interventions, often going well beyond what could be justified by the evidence presented.
- When the study design was less than ideal, often for good reasons, many studies did not discuss, let alone control for, the potential bias that may have resulted.

### Conclusions

Thus, our systematic review supports our contention that there are very few economic evaluation studies in the social welfare field outside health care, at least in published journals. Most of the studies we found are in areas that span the health care/social care boundary and most of these were carried out in the US. The quality of studies is also very variable.

If economic evaluation were to follow a similar pattern to the health care sector, then we would expect to see a rapid rise in the number of studies over time and a gradual reduction in the share of US studies. But, as yet, there is little evidence that the number of economic evaluations in the social welfare field is increasing.





# **Part II**

## **Conceptual issues**



# 4 Challenges for economic evaluation

## Why so little economic evaluation?

It is important to understand why there are currently few economic evaluation studies in many areas of social welfare, so that these obstacles can be addressed. There are several possible explanations (see Figure 6):

- Many social welfare interventions are very complex, which makes it difficult to carry out a systematic assessment of costs and outcomes.
- There may be resistance to the research methods favoured by economists from other researchers or practitioners working in this field.
- Rigorous economic evaluation can be resource-intensive if it requires new data to be collected and analysed over a long period of time.
- There may be a lack of demand from policy-makers and other decision-makers for this type of evaluation. The number of economic evaluations in the health care field is driven by regulatory and institutional structures, including, most recently, the role of the National Institute for Clinical Excellence in the assessment of new pharmaceutical and clinical interventions in England and Wales. The same pressures do not currently exist in the rest of social welfare. In other cases, the need for economic evaluation is recognised, but decision-makers are not always very discriminating (Knapp, 1999). They may have inflated expectations of economic evaluation, but lack the expertise to judge the quality of what they commission.

This chapter focuses on the first bullet point, examining the complexity of many social welfare interventions and the implications for evaluation. The next chapter addresses the second bullet point, by exploring alternative perspectives on evaluation and identifying a number of ways in which the conventional approach taken by economists (and other quantitative evaluators) differs from other approaches to evaluation. The third and fourth bullet points are discussed in Chapter 10, alongside other constraints on evaluation.

## Nature of programmes

The characteristics of many social welfare interventions make them hard to evaluate using conventional approaches to economic evaluation. Few people would argue against the view that the best way to evaluate the effectiveness of a new drug is to conduct a randomised controlled trial. So, what are the characteristics of drug-based interventions that lend themselves to a systematic and quantitative approach to evaluation, as favoured by most economists, and how do these differ from the characteristics of other health care interventions and other social welfare interventions? We use one of our case studies, the Safe In The City (SITC) programme (see Box 2), to illustrate some of these points. Although all interventions are different, the same issues are relevant to evaluation in many areas of social welfare, including more complex health care interventions. We also consider some of the practical implications for conducting economic evaluation in the social welfare field.

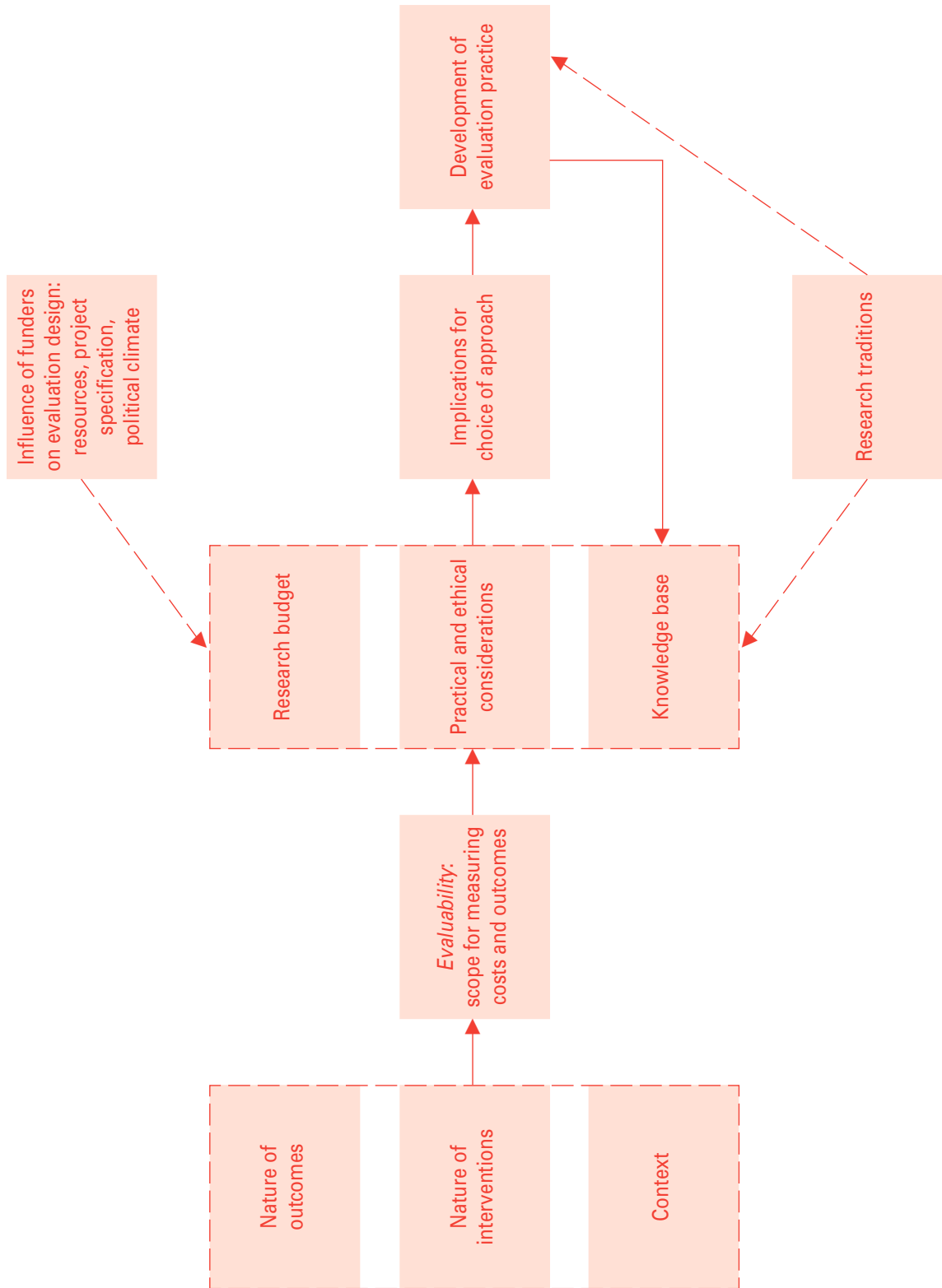


Figure 6 Scope for evaluation of social welfare interventions

### Complexity of outcomes

For a drug-based intervention, outcomes are often relatively easy to quantify, particularly for conditions that are physical (as opposed to psychological) and localised. For example, the effectiveness of certain classes of drug can be judged by their impact on mortality or disease-specific morbidity (e.g. cholesterol levels, blood pressure), which are relatively easy to measure. Other health care interventions seek to have a more 'holistic' impact on conditions that affect many aspects of a person's life, such as mental health problems. These are much harder to measure and compare, since they are less amenable to objective measurement and have several dimensions to them, including psychological, social, as well as physical functioning (Nocon and Qureshi, 1996). The broader social welfare field is similar. The SITC programme, for example, seeks to reduce the risk of youth homelessness by improving clients' self-esteem, encouraging them to take responsibility for their lives, and helping them deal with conflict in the home and at school. These types of outcome are much harder to assess in quantitative terms.

### Complexity of interventions

Drug-based interventions use standardised ingredients delivered in precise amounts to each patient. Some health care interventions, such as surgery, ought to be fairly uniform, but may vary in practice, because of differences in clinical practice between consultants or hospitals, or differences in the skills or competence of those delivering the service. Other health care interventions, such as counselling, need to be more flexible in order to match the varying needs of individual patients

or clients. Most social welfare interventions are similar in this respect. SITC projects, for example, are tailored to the needs of individual clients. It is also a heterogeneous intervention, made up of 24 delivery projects co-ordinated into eight cluster schemes, all with slightly different objectives, under the general banner of youth homelessness prevention. Knowledge of the extent and nature of variation within programmes is essential to understand how this affects outcomes. If, for example, the intervention appears to be more effective for some clients than others, it is helpful to know whether this is because these clients were more receptive to the intervention or because the intervention was different. Equally, when an intervention consists of a package of measures, it is useful to know whether some elements of the package are having more of an impact than others.

In addition, SITC is a developmental programme, which has been substantially restructured since it was first launched three years ago. Individual projects, too, are adapting what they do as they go along. In common with many other social welfare interventions, this programme is a moving target for evaluators. Even if an evaluation study could show that the overall programme was effective, it would be difficult to define what exactly it was that was working and, therefore, how to replicate it elsewhere.

### User involvement

Patients are fairly passive in a drug-based intervention. Other than complying with the treatment (e.g. remembering to take the drug in the appropriate dosages), patients have limited influence over the effectiveness or otherwise of

the intervention. Similarly, patients undergoing emergency surgery are fairly passive; although they can influence the recovery period, they have little impact on the success or otherwise of the surgical procedure. Although all interventions involve a degree of patient involvement, it is probably fair to say that this is not generally of the same order of magnitude as for more diverse community interventions in health care and other areas of social welfare. In the case of SITC, the success of the programme depends crucially on how young people respond to the intervention. This is determined by a whole range of factors, including whether they see the intervention as being helpful or an interference, how well they get on with project workers and other young people on the programme, and whether the intervention is reinforced by their family or school.

The more actively involved patients or clients are in the intervention, the more likely it is that outcomes will be affected by their attitudes, behaviour and circumstances. This introduces another set of confounding variables that need to be taken into account in an evaluation and that are harder to categorise or quantify in the same way as standard classifications, such as age and gender. This in turn makes it harder to generalise the results of an evaluation. Without a good understanding of how the intervention is working and, in particular, how people are responding to it, it cannot be assumed that the same intervention would have a similar impact in other settings.

### **Perspective and scope**

The effects of drug-based interventions are generally confined to patients and the health service, so it is relatively easy to achieve a

comprehensive picture of their impact – what economists call a societal perspective. A comparison of community and hospital-based care for elderly people with mental health problems would, however, need to be much broader, including the health service, but also social services, voluntary services, informal care and costs borne by both the patient and their family. Similarly, the evaluation of social welfare interventions will often span several sectors, since many different agencies are involved in providing services to vulnerable groups targeted by these programmes. An evaluation of the SITC programme, for example, would need to consider potential knock-on effects on social services, schools, employment services and voluntary services, as well as young people and their families.

### **Unit of analysis**

Drug-based interventions, like most health care services, are directed at individual patients. But some health care interventions seek to improve health (and possibly non-health) outcomes by influencing the structural or socio-economic causes of ill health. For example, a health promotion intervention might seek to reduce smoking by creating an environment that deters smokers (e.g. smoke-free areas, advertising bans). These kinds of community-based interventions are also becoming more popular in other areas of social welfare. In one of our other case studies, the Gatsby scheme, the aim is to help regenerate local communities by supporting small community groups. Although the aim is to improve the quality of life of people living in these areas, the focus of this intervention is on the community groups, rather than individual residents. These types of

intervention are more difficult to evaluate, because outcomes need to be assessed at different levels, including the organisation or community.

### Timescale

The effects of many drug-based interventions are intended to be fairly rapid, so that outcomes can be evaluated over a relatively short period without missing important effects (although possible side-effects may need to be monitored over a longer period). Similarly, some social care interventions address 'acute' problems, such as the provision of meals for an older person whose informal carer needs some temporary respite. Other health and social care problems, however, are more enduring and, therefore, costs and outcomes may need to be monitored over a much longer period. In addition, the timing of any impact may be uncertain, in which case the evaluation period may need to be extended at an additional cost. This is a particular problem for prevention programmes, such as SITC, which have a long gestation period or where the effects accumulate gradually over time. Projects that are directed at children as young as 13 are seeking to reduce the risk of them becoming homeless when they are 16 to 18.

A very crude method of scoring the ease of evaluating different programmes (or their 'evaluability') is set out in Table 1 (see overleaf). This covers the issues discussed above, as well

as other common evaluation problems. Social welfare interventions, like health care interventions, cover the whole spectrum from simple to complex. However, they typically have more in common with the treatment of complex and enduring health problems than drug-based interventions. Interventions are characterised by a high degree of user involvement, significant variability across recipients, complex and long-term outcomes and multiple agency involvement. Any one of these problems is likely to be encountered in most evaluation studies whatever the policy area. But, in the case of social welfare programmes and more complex health care interventions, it is likely that an evaluation will have to deal with several of them within the same study, which compounds the problem. For example, outcomes that are qualitative by nature can often be assessed using quantitative outcome scales. But, if on top of this, the effects are expected to be small and outcomes are influenced by many external factors, then it may be very difficult to measure accurately the impact of a particular programme.

Although these complexities should not preclude the achievement of a good quality economic evaluation, they do add significant difficulties and, as a result, evaluation studies will often fall short of the 'ideal'. Practical guidance on addressing these issues is provided in Part III of this report.

## Making the most of it

---

**Table 1 Criteria for assessing the evaluability of social welfare interventions**

---

<i>Nature of outcomes</i>	Well defined objectives or not?	Are the objectives of the programme not very clearly defined?
	Short or long-term	Is it difficult to assess final outcomes within the timeframe of the evaluation?
	Unique or multiple	Is it inappropriate to express the effects of the programme as a single or primary outcome?
	Quantitative or qualitative	Are outcomes qualitative by nature or difficult to measure for other reasons?
<i>Nature of programme</i>	Heterogeneity	Is programme a 'mixed bag' of different interventions or projects?
	Implementation	Is implementation likely to vary significantly within the programme?
	Unit of analysis	Is the programme directed at areas or communities?
<i>Context</i>	Scale effect	Is the potential impact on outcomes likely to be small relative to the scale of the problem?
	External influences	Is the potential impact on outcomes likely to be small relative to other 'external' influences on outcomes?
	Active or passive clients	Is the effectiveness of the programme very sensitive to how clients choose to respond to it?

---

Source: Based on Sefton (2000).

---



# 5 Alternative perspectives on evaluation

## Different schools of thought

Another reason why economic evaluation may have been slow to establish itself in some areas of social welfare is that most economists have not been very good at engaging with evaluators from other disciplines. The rest of this chapter discusses in broad terms some of the differences between different schools of thought on evaluation. While this discussion is highly simplified and presents two ends of a spectrum, it does serve to highlight a number of ways in which conventional approaches to economic evaluation might need adapting in a social welfare context.

## Summative versus formative

The focus of most economic evaluation is on testing well-defined interventions that are expected to remain fairly stable over time. If the intervention were to change significantly during the course of the evaluation, then, even if the results of the evaluation were promising, it would be difficult to say exactly what had been effective or cost-effective. Defining an intervention means specifying what services are to be provided, by whom, to whom and in what quantities. The intervention being evaluated should, ideally, be standardised with clearly written guidelines and checks to ensure that these are being adhered to in practice. This does not mean that clients should receive the same service regardless of their circumstances, but any variations should be systematic. Whereas economic evaluation tends to be summative, other approaches to evaluation are generally formative. The expectation is that most interventions will evolve over time in response to experience of operating the programme, changing circumstances, or simply the process

of implementing a programme, especially where different agencies are involved. Evaluators are interested in how programmes change over time and, in the case of action research, are actively involved in helping to shape this process (Greenwood, 1998).

## Quantitative versus qualitative

Economic evaluation is nearly always quantitative by nature, partly because this is the way economists are taught, but also because it is much easier to compare quantified costs and outcomes. Where outcomes are not naturally amenable to measurement, then outcome scales or proxy measures can be used. Study design is geared towards accurately quantifying the impact of a programme on costs and specified outcomes. Qualitative researchers, on the other hand, are more sceptical about numbers, which they would argue may provide a crude, and sometimes misleading, impression of a programme's impact. Numbers are not very good at capturing the richness and diversity of people's experiences and there is a danger of focusing only on those factors that can be measured easily. With quantitative analysis, it is also easy to ignore the context in which a programme is delivered, which may have a substantial influence on outcomes. Qualitative approaches are able to tell stories, to get underneath the surface and provide a better understanding of how a programme really affects an individual or institution.

## What versus how/why

Randomised controlled trials and other experimental approaches to evaluation are the basis for many economic evaluation studies, especially in health and social care. These focus

on the overall relationship between inputs and outcomes, often with little attention paid to how these outcomes are generated. Qualitative researchers are more interested in understanding how a programme works and their methods are more suited to addressing these types of question. They are not seen as distinct questions, but as part of understanding the programme's impact. There is a growing interest in theory-based approaches to evaluation, which put a greater emphasis on understanding the processes or mechanisms by which programmes work, as well as outcomes (e.g. Pawson and Tilley, 1997). This, however, is not an integral part of most economic evaluation.

### **Positivism versus constructivism**

Economic evaluation is aligned with a positivist view of the world. It is assumed that there are 'right answers' and it is just a matter of using the right method(s) to get to them. This is reflected in their approach to evaluation, including a strong preference for evaluation studies that are carried out at arm's length from those responsible for designing or implementing a programme so that evaluators can be as objective as possible. There is a reluctance to involve stakeholders in case this should influence the results. This also explains the concern with potential bias in the results (i.e. anything that might give a distorted picture of the 'true' impact). At the other extreme are constructivists who argue that there is no objective truth (e.g. Guba and Lincoln, 1989). Evaluating a programme is, according to this view, more like commenting on a painting or a book than measuring the height or length of something; in other words, truth or reality is in

the eye of the beholder. The role of the evaluator in this situation is to seek an agreed or negotiated consensus between different perspectives or, more realistically, to listen to different viewpoints and attempt to synthesise these (Kushner, 1996). Evaluations of this kind are a much more participative activity.

### **Individualism versus institutionalism**

One criticism of conventional economic approaches, including from some economists, is that they focus too narrowly on the outcomes for individuals and ignore changes to the broader socio-economic environment (for example, community values, infrastructure support and social capital). Proponents of this view argue that we need to move away from a purely individualistic (or utilitarian) framework to one that recognises that there is a complex and interactive relationship between individuals and their environment that helps to shape their behaviour. Thus, conventional approaches to economic evaluation will be less suited to evaluating programmes that have the community, rather than the individual, as the focus of interest – for example, a school-based intervention that seeks to reduce bullying by improving the school environment (e.g. by reinforcing positive norms). The impact of such programmes will take longer because time needs to be invested in creating the initial conditions for change (but it should also last longer); therefore, evaluations need to have a longer timeframe if the full benefits are to be included. Successful programmes will also leave the community better able to deal with its own problems in future and help to increase the capacity of the community to respond to future initiatives from outside – both important

outcomes that may not be captured in an individually focused evaluation. Institutional economists (and others) believe that an evaluation of a community-based intervention should seek to assess changes in the community, as well as outcomes for individuals, and should explore the (two-way) relationship between them. Otherwise, evaluation studies will be systematically biased against these kinds of interventions (Shiell and Hawe, 1996).

### Strengths and limitations of conventional approaches

The differences between these different perspectives, as we have described them, are exaggerated, but real. However, we believe that both perspectives have a place in the evaluation of social welfare interventions. The advantages of more quantitative approaches to evaluation are as follows:

- Evidence from experimental evaluations seems to hold more sway with policy-makers.
- They are more systematic, ensuring greater consistency of approach and facilitating comparisons across programmes.
- They force programmes to be very clear and specific about their goals, through having to express them in the form of quantifiable outcomes.
- It is easier to control for the effects of confounding factors.
- Quantitative results are easier to present concisely.

But, quantitative approaches to evaluation also have limitations, which are probably greater the more complex the intervention:

- Important outcomes may be difficult to measure, so they may be omitted or measured inappropriately.
- There may be practical and ethical obstacles to carrying out experimental studies.
- Evidence is only available at the end of the programme (or at pre-specified interim analysis points), which may be too late to make changes to the current programme and so is less likely to inform good practice in the short term.
- It imposes a fairly rigid framework on evaluation, so it may be less suited to evaluating programmes that are evolving rapidly over time.
- They do not tell you very much about how the programme works and why it is effective (or not).
- They may undervalue the importance of community outcomes.

In some cases, these potential shortcomings can be overcome through careful design or addressed by other components of a broader evaluation strategy. But, economists are also challenged to think more fundamentally about their approach to evaluation, in particular the need to consider being more flexible and eclectic in our choice of methodology, even if this does not always fit easily within a positivist or individualistic view of the world (Mannion and Small, 1999). Our response to this challenge is

reflected in the approach to economic evaluation outlined in the remainder of this report. This attempts to take on board some of the lessons from other perspectives without sacrificing the underlying strengths of more conventional approaches to economic evaluation.

### Possible framework for evaluating complex interventions

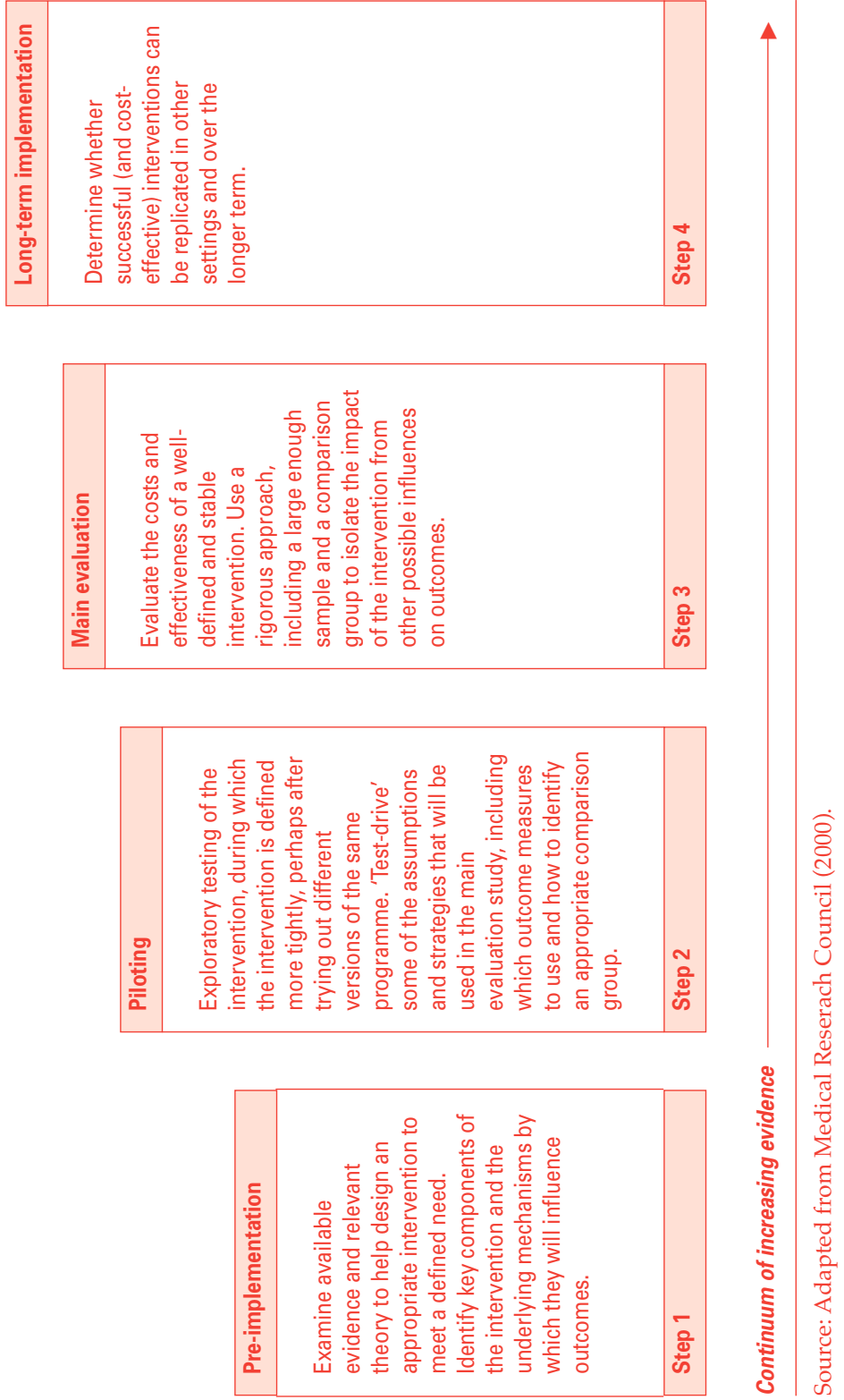
This section presents a useful framework for thinking about the evaluation of complex interventions, which is adapted from the different stages involved in evaluating new drugs (Medical Research Council, 2000). We are not suggesting that social welfare interventions should be tested in the same way as drugs or that all evaluation should be leading towards a large-scale controlled trial. But, a phased approach to evaluation is a helpful way of combining different evaluation activities into an overall strategy aimed at generating evidence on the effectiveness and cost-effectiveness of an intervention. We could equally have started with similar frameworks developed in other areas of social welfare (see, for example, the ‘five-tiered approach to evaluation’ described in Pecora *et al.*, 1995).

Figure 7 sets out the different steps involved in providing evidence to support a new intervention. The purpose and methods of evaluation will be different at each step, but each one informs the one that follows. At least some of the differences between the various perspectives on evaluation can be reconciled within this framework. In the early stages of evaluation, there is a clear role for formative approaches to evaluation in defining the

intervention, examining whether it is working as intended and helping to refine the programme into something that can be evaluated more rigorously. Quantitative methods have a more important role towards the latter stages of the evaluation process in terms of providing ‘scientific’ evidence on the effectiveness and cost-effectiveness of the programme. However, these may need to be complemented by the use of qualitative research methods in order to capture those outcomes (or aspects of outcomes) that are less amenable to measurement. Finally, a better understanding of how and why the intervention works would help to assess how far the results of the main evaluation study can be generalised to other settings.

While these phases are clearly delineated in the context of a drug trial, the process is likely to be much more ‘fuzzy’ in the rest of social welfare, which does not have the same institutional structures for testing interventions. But, it is still useful for policy-makers to consider where they think they are along this continuum and for evaluators to consider what techniques are appropriate at each stage and how these might feed into subsequent stages of evaluation. Table 2 lists some of the evaluation questions that may be relevant at different stages of an evaluation strategy, using the example of one of our case studies. Economic evaluation appears to come in only towards the latter stages of this evaluation strategy, but, in fact, many of the building blocks for economic evaluation should be put in place at an earlier stage.

Figure 7 Framework for the evaluation of complex interventions



**Table 2 Phased approach to evaluation (Safe In The City case study)**

Purpose	Specific evaluation questions
<b>Step 1: Pre-implementation</b>	
<ul style="list-style-type: none"> <li>• Documenting the need for a particular programme</li> <li>• Setting out basic characteristics of programme</li> <li>• Demonstrating fit between identified needs and programme</li> </ul>	<ul style="list-style-type: none"> <li>• What are the risk factors associated with youth homelessness? How prevalent are risk factors in the areas covered by Safe In The City? What services are already available in these areas and what are the gaps in the system?</li> <li>• What services are to be provided by the programme and whom are they targeted at?</li> <li>• What is the theory behind these services, i.e. on what basis have they been selected and how are they expected to have an impact on desired outcomes?</li> </ul>
<b>Step 2: Piloting</b>	
<ul style="list-style-type: none"> <li>• Clarifying projects' aims and objectives</li> <li>• Testing how the programme is operating in practice and seeking to make improvements</li> <li>• Developing an evaluation strategy for the next step</li> </ul>	<ul style="list-style-type: none"> <li>• How is the concept of homelessness prevention interpreted by the different clusters of agencies involved in the programme? Do clusters have clearly stated objectives and are these consistent with the programme's overall objectives?</li> <li>• How are specific elements of the programme being implemented? What is emerging as 'best practice' from different projects?</li> <li>• What costs and outcomes should be assessed and how? What would be a suitable comparison group and how might they be recruited?</li> </ul>
<b>Step 3: Main evaluation</b>	
<ul style="list-style-type: none"> <li>• Assessing whether individual projects are meeting their own objectives</li> <li>• Evaluating the effectiveness and cost-effectiveness of programme, compared with other programmes that have similar objectives</li> </ul>	<ul style="list-style-type: none"> <li>• Are individual projects meeting their objectives? How do their clients feel about the services provided?</li> <li>• What is the net impact of the programme, allowing for other influences on outcomes? Is prevention cost-effective compared with alternative approaches (e.g. 'crisis' interventions)?</li> </ul>
<b>Step 4: Long-term implementation</b>	
<ul style="list-style-type: none"> <li>• Suggesting programme or project models worthy of replication</li> </ul>	<ul style="list-style-type: none"> <li>• What types of intervention appear to work best, for whom and in what circumstances?</li> </ul>

# **Part III**

## **Practical application**





# 6 Clarifying purpose and objectives

## Key aspects of evaluation

Economic evaluation, as we have defined it, involves a systematic assessment of the costs and outcomes of an intervention. The key components of economic evaluation are described in Figure 8.

This chapter discusses the initial or preparatory stages of evaluation – the top two boxes in Figure 8. Chapters 7 to 10 discuss other aspects of economic evaluation, including the assessment of costs and outcomes, study design, analysis and interpretation of results, and managing the evaluation process.

## Assessment criteria

Before considering these issues in turn, it is worth considering the criteria against which the design of an evaluation should be assessed:

- *Validity*: this is the extent to which the evaluation provides an accurate assessment of outcomes and costs. Economists (and other quantitative evaluators) usually focus on a statistical definition of validity, i.e. the extent to which the method used provides an unbiased estimate of outcomes and costs (e.g. Moser and Kalton, 1993). For qualitative researchers, validity depends on how faithfully a study conveys people's experiences of an intervention (e.g. National Centre for Social Research, 2000).
- *Generalisability*: this is about the wider applicability of a study's results in other settings or for other client groups. Quantitative evaluators focus on how statistically representative their study sample is of the target group as a whole

(e.g. Moser and Kalton, 1993). A broader definition would include other types of evidence, including knowledge about why certain interventions work in certain circumstances and for certain groups of people.

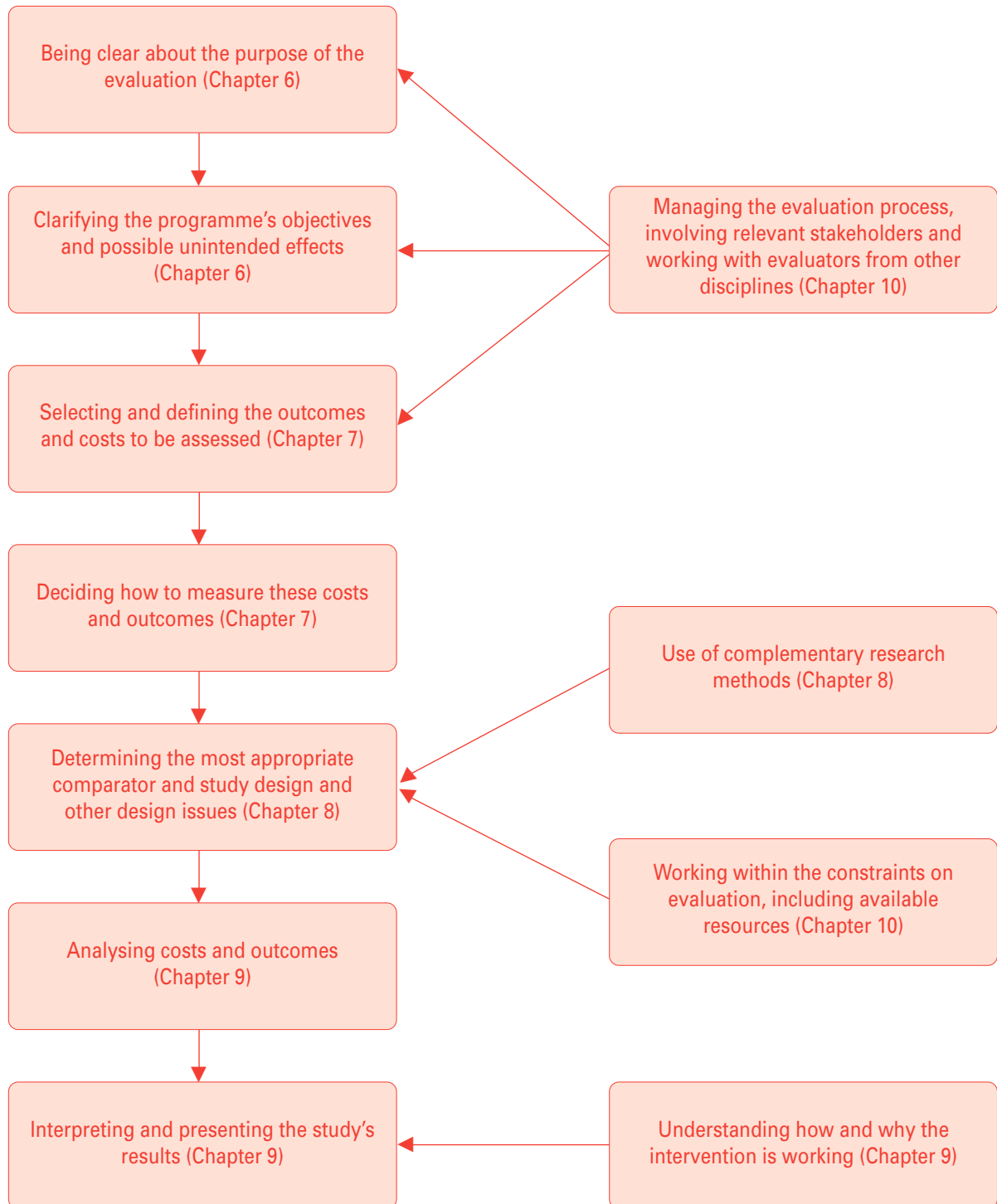
- *Relevance*: this is about how useful the results are to decision-makers. Most economic evaluation studies are designed to answer a specific question: whether the intervention being evaluated is an efficient use of resources, compared with alternative ways of using these resources. But, decision-makers may be interested in a much wider set of evaluation questions, such as how to improve a programme.
- *Feasibility*: this is the extent to which the proposed evaluation strategy can be implemented in practice, given the constraints on evaluators.

All of these criteria are important. For example, there is little point in obtaining an exact estimate of a programme's impact in a particular setting if this is untypical of the settings in which this programme would normally be implemented. Similarly, there is little point in designing a very sophisticated evaluation if it would be too costly or would not address those questions of greatest interest to decision-makers.

## Being clear about the purpose of evaluation

Before designing an evaluation strategy, it is important to be clear about the purpose of the evaluation. Eventually, the aim is to identify a well-specified intervention that generates desirable outcomes, offers value for money

Figure 8 Key stages of economic evaluation



compared to other interventions with similar objectives and can be replicated elsewhere. However, there may be other priorities for evaluation, especially early on in a programme's development. These include:

- documenting what the programme is doing and where expenditure is going, to ensure accountability to those funding the programme
- checking whether the programme is operating as originally envisaged, especially if there are several agencies involved
- looking for ways to develop a programme, in order to iron out any teething problems and improve its effectiveness
- monitoring progress towards meeting programme objectives
- providing 'ammunition' for public relations purposes, to help 'sell' the programme to funders and policy-makers
- motivating project workers and clients, by helping them to see the progress they are making.

Much evaluation in the field of social welfare is focused on these sorts of issues (Newburn, 2001). While economists may have less to contribute to these more formative types of evaluation, there are several implications for economic evaluation. First, some of these activities may feed into the economic evaluation, so it is important for economists to be involved early on. For example, the information systems set up for monitoring or

accountability purposes may also provide the data needed for an economic evaluation. Similarly, the process of documenting a programme and identifying its key components may inform the choice of outcomes.

Second, it is important to recognise that there may be a conflict between the different roles of evaluation. In our Safe In The City case study, individual projects saw evaluation as a useful way of demonstrating to clients the progress they were making. But if evaluation is being used in this way, there is likely to be a bias in favour of recording positive changes in outcomes. It may, therefore, be necessary to keep this process separate from the collection of data for economic evaluation. The pan-London co-ordinating organisation had a different perspective. Although it has since adopted a different approach to evaluation, initially it felt that it had to focus on measuring the impact on youth homelessness, because this was the programme's 'headline' objective, despite some of the problems in attempting to do this (see Box 4 in Chapter 7).

Third, it is important that economists acknowledge these other priorities and recognise that they need to be addressed alongside economic evaluation. Thus, economic evaluation should be seen as just one part of an overall evaluation strategy. One problem is knowing when to carry out an economic evaluation. It is probably not appropriate to conduct a controlled experiment of a programme that is in its very early stages of development and is being continually modified. At the same time, it is important to assess the cost-effectiveness of new programmes as early as possible to see if they are worth continuing or rolling out.

### Clarifying programme objectives

Most evaluation, including economic evaluation, is goal-based. It is assumed that all programmes should have clearly defined objectives that are set out in advance and against which their performance can be assessed (e.g. HM Treasury, 1997). But, this is often not the case in the social welfare field:

- Objectives may not be very explicit or may be stated only in general terms.
- Objectives may change over time, often subtly, as the programme evolves.
- The ‘operative’ objectives, as reflected in the way the programme operates in practice, may differ from ‘stated’ objectives, especially if there are different agencies involved in delivering the programme.
- There may be several, possibly conflicting, objectives, perhaps because there are different stakeholders involved, and it may not be clear what priority is attached to different objectives.
- Some programmes are purposively flexible and developmental. Often, the rationale is that programmes need to be responsive to local needs, so local agencies may be invited to come up with local solutions to local problems. The funder is not prescriptive about the design of individual projects; indeed, they may wish to encourage diversity in order to promote policy innovation (Martin and Sanderson, 2000).

Some economists would see ill-defined objectives as a reflection of poor policy design, rather than the natural consequence of a ‘messy’ policy-making process. ‘Goal-focused evaluation’, on the other hand, recognises that goals or objectives are rarely as clear-cut as many evaluators would like. Clarification of objectives is made a central focus of the evaluation (i.e. an end in itself), rather than just the starting point (Peled and Spiro, 1998). This is an iterative and interactive process whereby declared or stated objectives are identified, compared with operative objectives, discussed with stakeholders and then revised if appropriate. Only then does evaluation proceed to an assessment of whether these objectives are being achieved.

None of this denies the importance of having clearly defined objectives as the ‘cornerstone’ for economic evaluation, but there are implications for the way economists think about objectives. For many social welfare interventions, objectives are not a ‘given’; indeed, the process of clarifying objectives is a fundamental and substantive element of an evaluation study. While responsibility for articulating the aims and objectives of a programme rests ultimately with those in charge of its implementation, evaluators can and ought to facilitate this process (Shiell and Hawe, 1996). Box 3 discusses these issues for two of our case studies.

### Box 3 Clarifying objectives

#### *Safe In The City case study*

One of the complications of this programme in terms of clarifying objectives at an early stage in the development of the new 'cluster' programme is that it was composed of 24 different projects, all with different objectives. This is not surprising given that delivery of the programme was through independent agencies, and the co-ordinating mechanisms designed to give coherence were not yet in place. We visited nine projects, covering the main types of project (employment and skills, personal development, and family support) to find out what they were trying to achieve and, therefore, what sort of outcomes we should be seeking to evaluate.

Interestingly, project workers did not talk very much about preventing youth homelessness. Nor did it enter into their own assessments of their project's effectiveness, which concentrated on shorter-term objectives, such as helping young people to take more responsibility for their lives and encouraging families to manage their own problems.

At first glance, the aims of different projects seemed to vary considerably, but discussions with project workers suggested that there was in fact significant overlap between what the different projects were seeking to do, though often by different means. For example, the employment and skills projects recognised that personal development (e.g. building self-esteem) was a very important, outcome of their projects, as well as a means to achieving other outcomes (e.g. better employment prospects). Thus, differences between projects were more to do with emphasis than substance. The multi-agency structures and protocols introduced more recently by Safe In The City add coherence to the programme's objectives.

#### *Fuel poverty case study*

The Home Energy Efficiency Scheme (HEES) aims to reduce the number of households living in fuel poverty. But, some ambiguities remain in the definition of fuel poverty and, therefore, the precise objectives of HEES. The most common definition of fuel poverty is couched in terms of 'affordability' (i.e. not being able to heat your home adequately without spending more than 10 per cent of household income). There is an important distinction between the amount a household would need to spend to heat their home satisfactorily and the amount they actually spend. Some households are fuel poor, but keep their homes warm by spending a very high proportion of their net income on fuel, while others could afford to heat their homes (without spending more than 10 per cent of their income), but do not do so. Thus, fuel poverty is not synonymous with cold homes, although there is a considerable overlap between the two. The Government's target focuses on the numbers of vulnerable households in fuel poverty. But, they and other agencies are clearly concerned about the numbers of people living in cold homes (some of whom are not fuel poor), because of the adverse health effects and, in the most extreme cases, cold winter deaths.

*continued overleaf*

In addition, HEES has potential environmental benefits, primarily in terms of reduced CO<sub>2</sub> emissions. While these were not mentioned explicitly in the Government's consultation paper on HEES (Department of the Environment, Transport and the Regions, 2000), energy efficiency is an important element in the Government's climate change strategy. Energy efficiency improvements can contribute to both social and environmental objectives: households may be able to heat their home better and at the same time cut down on their fuel consumption or switch to more environmentally friendly sources of energy. However, there is a potential conflict between the two objectives, since the greatest energy savings may be had from better-off households, who have the highest fuel consumption. Hence, the appropriate targeting of the programme will depend on the weight that is attached to different policy objectives.

An economic evaluation should consider all the scheme's objectives and the potential conflict between them. Even if environmental benefits were not one of HEES's stated objectives, they would still need to be taken into account in a full economic evaluation of the scheme.

# 7 Assessing outcomes and costs

## Choosing the right outcomes

The main determinant of outcomes should be the programme's objectives (see Chapter 6). But, there are two important respects in which economic evaluation diverges from purely goal-based evaluation. As already noted, economists prefer, if possible, to look at final outcomes (e.g. quality of life measures), whereas a programme's objectives may be expressed in other terms. For example, the objectives of an estate management project may focus on an intermediate outcome (e.g. faster repairs) or even a process measure (e.g. consultation procedures with tenants), rather than the final outcome, which would be improvements in tenants' quality of life.

Second, economic evaluation aims to be comprehensive, so it is important to consider other, perhaps unintended, outcomes, which may not be referred to in the objectives. This includes the impact on parties other than the main target group. This is already well recognised in some areas of social welfare. For example, it is common practice in the field of crime prevention to consider the knock-on effects on adjacent areas (e.g. Welsh and Farrington, 1999). Similarly, it is standard practice in evaluations of welfare-to-work programmes to consider displacement and substitution effects, i.e. that some participants may secure employment at the expense of other job seekers (Gardiner, 1997). In the social care field, there is a long history of considering the impact on carers, as well as clients (e.g. Davies and Challis, 1986; Knapp *et al.*, 1992).

## Evaluability of outcomes

However, practical considerations also need to be taken into account in selecting outcomes. It may not always be appropriate to choose the 'obvious'

outcome, because this may be difficult to measure accurately or because changes are expected only in the longer term, beyond a realistic timeframe for evaluation. In addition, the effects of the intervention may be difficult to disentangle from other external influences, especially if the expected impact is relatively small. For these reasons, it may be appropriate to select other outcomes, for example intermediate or shorter-term outcomes that are more directly affected by the intervention. Some examples from our case studies are discussed in Box 4.

## Multiple outcomes

In the social welfare field, it will usually be appropriate to include several outcomes in the analysis for the following reasons:

- Some programmes have multiple objectives.
- The objective, such as improving older people's well-being, may be multi-dimensional, so it may help to separate this into its different components (e.g. mobility, autonomy, appearance, social).
- Interventions may impact at different levels: it may be important to evaluate the impact on the community or organisation in which the intervention is set, as well as on individuals, in order to capture the full impact of a programme.
- Having several outcomes is useful for validation purposes, especially if there are some doubts about the reliability of outcome measures. For similar reasons, it is sometimes useful to have different perspectives on the same outcome; if, for example, the outcome is young persons'

behaviour, then it may be worth asking the views of parents and their teachers or social workers, as well as the young person's own views.

- Multiple outcomes can be used to evaluate potential knock-on or spillover effects, ensuring the evaluation is more comprehensive.

- Intermediate or process outcomes can be monitored in order to help interpret changes in the principal outcome, especially if there are known (or theorised) links between them.

Box 5 shows how multiple outcomes have been used in a range of published studies.

### Box 4 Choosing outcomes

#### *Gatsby case study*

This case study illustrates the process of selecting an appropriate outcome for a complex intervention. Half the budget is spent on residential training for community groups, which is seen to be worthwhile in its own right, as well as being a gateway to the grants scheme. Grant-funded projects are part of a learning process for groups, which is why the scheme is targeted at relatively new groups and why ongoing support is provided to groups. Thus, the scheme has a strong element of 'capacity-building' that needs to be reflected in the choice of outcome.

We considered three levels at which this scheme could be evaluated:

- social and economic impacts on the *local community*
- success of *individual projects* funded by the scheme
- impact on the effectiveness of *community groups*.

Of these, the third option was thought to be the most promising. The impact on community groups is a direct one, so it should be easier (though still not easy) to identify than any impact on the wider community, which may be too small to measure accurately or to disentangle from other factors. This outcome would also capture the two key elements of the scheme; training has a direct impact on the skills, commitment and vision of groups, while grants help groups to put their ideas into practice and feed indirectly into the skills and knowledge of the groups through the experience of running projects (i.e. learning by doing). In addition, this approach is consistent with the principles of community development, by recognising that process matters (i.e. the focus is on helping local communities to help themselves through building the capacity of community groups).

#### *Safe In The City case study*

This case study is a good example of why the choice of outcome may not be as straightforward as first appears. The 'obvious' outcome would be reductions in homelessness among the programme's clients, which is their 'headline' objective. However, there are several reasons



why this may not be the most appropriate outcome:

- The timing of any impact on homelessness is both long-term and uncertain, so there is a danger that a study focusing on this outcome would be unable to identify any significant change over the likely evaluation period (i.e. two to three years at most).
- Even if the evaluation were extended over a sufficiently long period, it would be very difficult to keep track of former clients. The validity of the study could be significantly compromised by sample attrition, especially if, as seems likely, those with less stable housing outcomes are hardest to track.
- Even if these problems could be overcome, the impact on the numbers of homeless young people may be too small to detect. This does not mean the programme is not worthwhile, because a small impact may be sufficient to justify the programme, given the high and long-term costs of youth homelessness. But, it does mean that even quite a large-scale evaluation study (involving, say, 300–500 respondents) might not be sufficient to identify a significant impact.
- Last, but not least, we think it would be unwise to focus all the attention on a single outcome measure, given the multiplicity of objectives cited by the projects we visited.

For these reasons, we think it would be too risky to focus on homelessness as the only, or even the primary, outcome measure. Instead, we recommended looking at a number of shorter-term outcomes, including reduced emotional and behavioural problems, reduced family conflict and improvements in young people's quality of life, although it would make sense to monitor housing outcomes, as well.

### *Revolving Doors Agency*

A slightly different situation is when the intervention being evaluated is one of a number of interventions impacting on the same target group. For example, the Revolving Doors Agency (RDA) seeks to improve access to mental health services for people with chaotic lifestyles who have been in contact with the police (Revolving Doors Agency, 2000). The final outcome is an improvement in the mental health of its clients, but this also depends on the quality of local health, social and housing services, which is largely outside the Agency's control. Arguably, the success of RDA should be judged on its effectiveness in encouraging greater or more appropriate service use by its clients, which it has a direct impact on, rather than changes in clients' mental health, which it influences only indirectly.

However, there is an additional complication. As well as encouraging their clients to make more appropriate use of public services, Link Workers have also taken on a mentoring role, forming close relationships with some of their clients. As a result, the scheme may have a direct impact on the mental health of clients, which would be omitted if the evaluation were only to examine the impact on clients' use of services.

### Box 5 Multiple outcomes

#### Example 1

This is an experimental study of school social work at a primary and linked secondary school. Data were collected on a range of indicators, including truancy, self-reported theft, bullying, drug use, vandalism and school exclusions due to behavioural disruption. This range of outcome measures is appropriate, because the project was aiming to have an impact on all these variables. But, the fact that they were able to show positive outcomes on six out of seven indicators also helps strengthen the validity of the results. It is much more convincing than if they had selected just one of these indicators and shown an improvement in this.

(Source: Bagley and Pritchard, 1998)

#### Example 2

This economic evaluation compares the Darlington model, a scheme to provide intensive and better coordinated support for frail elderly people in a community setting, with long-term hospital care. This type of intervention is likely to have an impact on many different aspects of an older person's quality of life. Hence, it was important to include a wide range of outcomes, including both psychological indicators (e.g. general life satisfaction, depression, loneliness) and behavioural indicators (e.g. physical disability, apathy, communication difficulties). The study also monitored mortality rates for both groups of patients.

(Source: Challis *et al.*, 1995)

#### Example 3

This evaluation of a screening programme for aboriginal infants looked beyond direct measures of health outcome at other ways in which the intervention might affect the community in which it was set, including positive changes in this community's perception of health care services (e.g. greater willingness to visit a doctor). These indirect outcomes were seen to be valued outcomes in themselves and should improve the community's capacity to translate future investment in health services into health gains.

(Source: Jan, 1998)

#### Example 4

This is an evaluation of a UK welfare-to-work scheme (RESTART), which provided advice and ongoing support to unemployed people. Although the main outcome measure was a reduction in benefit claimants, information on a range of variables was also collected, including the quality of jobs (using wages as a proxy) and process measures (using numbers of job applications as a proxy for job-search activity). These helped to interpret the results of the study and to explore various hypotheses about how the scheme was working (e.g. whether it was pushing people into lower quality jobs).

(Source: White and Lakey, 1992)

From an economist's point of view, multiple outcomes do not usually allow interventions to be ranked neatly in terms of their cost-effectiveness, unless one intervention is no more expensive and performs better on all the outcome measures (or is cheaper and performs no worse on any of the outcome measures). It is common practice in the health care field to specify a primary outcome in advance, which enables a cost-effectiveness analysis to be conducted. But, this is only appropriate if there are good reasons for giving more weight to one particular outcome.

### Defining outcomes

Before moving onto the measurement of outcomes, it may be necessary to define these outcomes more precisely. For example, what exactly is meant by 'better' or 'more secure' jobs in the case of a welfare-to-work programme? What precisely is meant by a better school environment in the case of a school-based social work intervention?

In some cases, this will follow from the programme's objectives, provided these have been clarified in sufficient detail. For example, a school-based social work programme may specify that its aims are to reduce bullying and the numbers of exclusions. In other cases, outcomes may need to be defined almost from scratch. In our Gatsby case study, we had to define what an effective community group was, before we could set about trying to assess groups against these criteria. The process of defining outcomes is likely to involve discussions with key stakeholders, in particular those responsible for setting the programme's objectives. We also reviewed the literature for evidence and theories about effective community organisations.

### Measuring outcomes

Some outcomes are more amenable to measurement than others. For example, the yardstick for evaluating welfare-to-work programmes in the UK is the net effect on employment, whilst, for crime prevention programmes, it is the reduction in different types of crime. But, even in these cases, outcome measurement may not be straightforward. For example, evaluations of welfare-to-work programmes should ideally take into account the quality of jobs (i.e. pay, job security), as well as the number of jobs created. There may also be measurement problems. This is a particular concern with crime statistics, where official statistics are known to be substantial under-estimates and, if asked directly, offenders have clear incentives to under-report. In addition, if outcome measures are being used to assess an organisation's performance, this may distort the results, especially where there is some discretion in the way statistics are recorded. (For example, police records may be less likely to record a crime as being racially motivated if the police are under pressure to cut this type of crime.) Even when 'hard' data are available, these may be indicators rather than outcome measures. The temptation, which must be avoided, is to use whatever data is available, even if this is not a very good proxy for the outcome that is being assessed.

Many outcomes are not naturally amenable to measurement. These are sometimes referred to as 'soft' outcomes (e.g. the effectiveness of a parent, the well-being of an older person, the behaviour of a young person). There seems to be a common perception among many working in the social welfare field, including other

evaluators, that these kinds of outcomes cannot be measured and so this is often used to reject quantitative approaches to evaluation. Even the term ‘soft’ implies that these outcomes are somehow less substantive than ‘hard’ outcomes. For example, SITC recognised that ‘soft’ outcomes, such as improvements in clients’ self-esteem were important and wanted these to be monitored; however, it was felt that policy-makers would mainly want to know about hard outcomes, such as reductions in youth homelessness, as this had been emphasised heavily in the past.

Our view is that many of these outcomes can be measured in some way (see below), allowing all sorts of interventions to be evaluated within a quantitative framework. However, it is necessary to be pragmatic about the measurement of very complex outcomes, which

may need to be assessed qualitatively (but should not simply be ignored).

### Established outcome scales

In some fields, such as psychology, there is a strong tradition of using quantitative scales to assess outcomes that are not naturally amenable to measurement, such as family functioning, feelings of depression, or self-esteem (e.g. Department of Health, 2000). These are usually based on questionnaires that can be scored on a numerical scale. Box 6 gives an example of one such scale, the Strengths and Difficulties Questionnaire, which is used in the foster care case study. Where outcome scales are not available, it may be possible to develop a new scale. Box 6 gives an example of an evaluation tool we put together as part of our Gatsby case study to help assess the effectiveness of

### Box 6 Use of outcome scales to measure ‘soft’ outcomes

#### Example 1

The Strengths and Difficulties Questionnaire (or SDQ) is a widely used and validated instrument for screening for emotional and behavioural problems in children and adolescents. There are different versions to be completed by parents, teachers and young people aged 11 to 16. The scale consists of 25 items or questions, which are scored on a scale of 0 to 2.

Respondents are asked to give answers on the basis of how they have been feeling over the last six months. The whole questionnaire takes around ten minutes to complete. The first ten questions are:

	Not true	Somewhat true	Certainly true
1 I try to be nice to people, I care about their feelings	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2 I get restless, I cannot sit still for long	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3 I get a lot of headaches, stomach aches or sickness	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4 I usually share with others (food, games, pens, etc.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5 I get very angry and often lose my temper	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

- |    |   |                          |                          |                          |
|----|---|--------------------------|--------------------------|--------------------------|
| 6  | I usually do as I am told                             | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 7  | I worry a lot   | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 8  | I am helpful if someone is hurt, upset or feeling ill | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 9  | I am constantly fidgeting or squirming                | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 10 | I have one good friend or more                        | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

**Example 2**

We developed our own questionnaire to assess the effectiveness of community groups, because there were no established outcome scales. This was based on a series of statements that group members were asked to agree or disagree with. The statements are mostly quotes from previous visits to other community groups. There are four sections, each with ten questions. Each section takes at most five minutes to complete. This section is about how well the team works together:

- |   | Strongly agree           | Agree                    | Neither agree nor disagree | Disagree                 | Strongly disagree        |
|---|--------------------------|--------------------------|----------------------------|--------------------------|--------------------------|
| 1 'There's a lot of infighting – sometimes you could cut the atmosphere with a knife' | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/>   | <input type="checkbox"/> | <input type="checkbox"/> |
| 2 'We're lucky to have such strong leadership'  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/>   | <input type="checkbox"/> | <input type="checkbox"/> |
| 3 'Some people are trying to play the same role – stepping on each other's feet'      | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/>   | <input type="checkbox"/> | <input type="checkbox"/> |
| 4 'Sometimes it's hard to see who's in charge'  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/>   | <input type="checkbox"/> | <input type="checkbox"/> |
| 5 'People are willing to put themselves out for the group'                            | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/>   | <input type="checkbox"/> | <input type="checkbox"/> |
| 6 'Any disagreements tend to be over quite minor things'                              | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/>   | <input type="checkbox"/> | <input type="checkbox"/> |
| 7 'Morale is quite low – people are thinking of packing it all in'                    | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/>   | <input type="checkbox"/> | <input type="checkbox"/> |
| 8 'We complement one another as a group – it works'                                   | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/>   | <input type="checkbox"/> | <input type="checkbox"/> |
| 9 'When the going gets tough, people don't want to know'                              | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/>   | <input type="checkbox"/> | <input type="checkbox"/> |
| 10 'It's really satisfying what we're doing as a group'                               | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/>   | <input type="checkbox"/> | <input type="checkbox"/> |

community groups (see Example 2). Volume II of this report provides more information on developing outcome scales and further examples of existing scales that may be useful in the social welfare field (Byford *et al.*, 2002, forthcoming).

If conducting a rigorous experimental evaluation, there should be a general presumption in favour of using established outcome scales, such as the Strengths and Difficulties Questionnaire, wherever possible. The scales have been tested in various ways, which helps to increase the validity of the results. This also increases the comparability between studies if the same outcome scales are used widely.

But, established scales need to be employed with some care. It is important for evaluators to ensure these scales are measuring the right thing. For example, some scales, such as the EuroQol, have been developed to measure health-related quality of life and focus more on physical than psychological or social functioning (EuroQol Group, 1990). Although they have been validated for use in the health care field, they will not necessarily be suitable for measuring change in a broader social welfare context. Lack of appropriate scales may hamper the measurement of outcomes in more complex areas of mental health and social care, although lack of awareness and understanding of suitable instruments seems to be the greatest constraint at present. In our Safe In The City case study, the organisation was concerned that established scales would miss out important aspects of outcomes, so they designed their own set of soft indicators in consultation with project workers. In practice, the best option would probably be to use a mix of both types of

measure, subject to limitations on the resources and time available for data collection. In the case of SITC, there was a particular concern not to overburden clients with too many questions, especially early on in the intervention.

Another problem is that many outcome scales have been developed for diagnostic purposes (e.g. to identify children with problem behaviours), rather than for evaluation purposes. Therefore, they may not be very sensitive to change (Pecora *et al.*, 1995). In addition, a scale may have been designed to pick up extreme cases, in which case it would not be so good at picking up changes lower down the scale.

### Impact on respondents

It is also important to consider the impact on respondents, especially more vulnerable clients. Although these concerns are sometimes exaggerated, researchers should ensure questions are culturally sensitive and are strengths-oriented (i.e. that they do not concentrate only on negative aspects). For example, the Strengths and Difficulties Questionnaire includes a section on pro-social behaviours, as well as sections on anti-social behaviours.

### Self-report scales

Self-reported outcomes are common in the social welfare field. They are appropriate if outcomes are subjective by nature (e.g. feelings of loneliness or depression) or if they are hard for the researchers to monitor directly (e.g. someone's drinking behaviour or incidents of domestic violence). They also take users' views into account, which are being given increasing prominence by funding bodies (e.g. Joseph

Rowntree Foundation, 2001b). Self-reported outcomes are less appropriate if respondents are not in a very good position to make these judgements, because they have incomplete information, or because it is hard for them to be objective about their own situation (e.g. about their own employability), or because there are disincentives to providing accurate information (e.g. on criminal behaviour). These different considerations need to be weighed up. Generally, it is harder for people to assess changes in outcomes than absolute levels. It is better, if possible, to measure outcomes at two or more points in time and compare than to ask individuals to assess changes over time.

### Validity and reliability

Confidence in the validity of outcome measures can be improved in several ways:

- using validated scales in preference to, or alongside, 'do-it-yourself' outcome measures
- monitoring several distinct, though related, outcomes to see if they tell a consistent story
- collecting data from different sources (e.g. from clients' partners or carers)
- using a combination of subjective measures (e.g. self-reported drinking behaviour) and objective measures (e.g. blood tests to measure alcohol levels)
- carrying out a qualitative assessment of outcomes, perhaps on a sub-sample of clients, alongside the use of quantitative outcome measures.

### Valuation of benefits

Cost-benefit analysis requires outcomes, as well as costs, to be valued in monetary terms. The advantage is that the benefits of an intervention can then be compared directly against the costs.

Some outcomes are relatively easy to value. For example, the benefits of a welfare-to-work programme can be valued by monitoring participants' earnings (which, in a competitive labour market, will equate to the economic value of their work). Similarly, the benefits of an energy efficiency scheme can be valued in part by estimating the value of potential fuel savings at market prices. However, many outcomes of social welfare interventions will not have a price attached to them, because they are non-marketed goods or services. How, for example, do you value the benefits of not being mugged or feeling safer at home?

Economists have developed sophisticated techniques for putting monetary values on some of these outcomes. Revealed preference techniques look at people's spending on related goods to see what this indicates about their willingness to pay for non-marketed goods and services. For example, they might look at how much people spend on burglar alarms or the premium they are willing to pay for houses in low-crime areas in order to estimate the benefits of greater security. An alternative approach is to ask people directly what they would be willing to spend for a positive outcome (or how much they would be willing to accept in compensation for a negative outcome). Stated preference techniques usually involve asking a representative sample of people to make hypothetical choices between two or more clearly defined scenarios. For example, people living in a high-crime area might be given a

description of what the area would be like with lower levels of crime and asked if they would be willing to pay £*x* more in local taxes to achieve this change. The advantage of stated preference techniques is that they are very flexible and can be used to estimate the value of almost anything.

However, valuation methods are controversial and rarely used as part of an economic evaluation. The validity of estimates has been questioned and end users may be reluctant to accept evaluations with monetised outcome measures (Hutton, 1992; Johannesson *et al.*, 1996). In addition, valuation methods often involve surveys of the general public, which is a time-consuming and expensive process if conducted properly.

In practice, in the social welfare field, we identified very few economic evaluations that made use of revealed or stated preference techniques to value outcomes in monetary terms. One study used broad estimates of the value of life to assess whether the benefits of parental leave (in terms of reduced child mortality) were likely to outweigh the costs (Ruhm, 2000). Another study put a value on the comfort benefits from energy efficiency improvements, by estimating how much of the potential savings households spent on additional heating (Clinch and Healy, 2001). We are also aware that the Home Office is exploring the use of these techniques to value the benefits of reducing crime (Healey, 2001).

But, most studies that were classified as cost-benefit analyses were using much cruder methods to monetise the benefits of programmes. For example, the benefits of reduced crime were typically valued by estimating potential savings to the criminal

justice system and, in some cases, the value of lost property (e.g. Ekblom *et al.*, 1996; Forrester *et al.*, 1990; Robertson *et al.*, 2001). These studies, which are really cost-savings analyses, are likely to under-estimate the ‘true’ benefits of crime prevention programmes, because they ignore the psychological costs to victims and to others affected by the fear of crime. However, they can still be very useful in making a case for a particular programme if they show that potential savings exceed costs.

### Scoring and weighting

Scoring or weighting is another method of combining different outcomes into a single index. Option appraisal is one such technique that has been used to assess proposals for capital spending in the health service (NHS Executive, 1994) and could be adapted for use in the social welfare field for appraisal or evaluation purposes (Holtermann, 1998). This approach involves listing all the criteria that are relevant to assessing a particular type of proposal, giving each option a numerical score on each criterion, giving each criterion a weight to reflect its relative importance and then multiplying the scores by the weights to give an overall score for each option.

The advantage of this approach is that it avoids the need to put a monetary value on outcomes, which many non-economists and some economists are wary about. Like monetary valuation, it enables direct comparisons to be made between competing schemes or projects (though only if they can be assessed against the same criteria). Although the actual scores or weights may be somewhat arbitrary, proponents of this approach would argue that it does at least ensure that spending decisions are made



on a more explicit and rational basis. These techniques can be strengthened by involving experts, including service users, in agreeing the weights or scores to be used and by exploring the sensitivity of the rankings to different choices.

### Assessment of costs

This section considers the definition of costs and some of the barriers to the accurate measurement of costs, including some specific challenges for the costing of social welfare interventions. More detailed guidance is provided in Volume II of this report (Byford *et al.*, 2002, forthcoming). The most common understanding of the term costs is the amount spent on providing certain goods or services. By contrast, economists are interested in opportunity costs. These are the benefits foregone by tying up resources in one particular use and so not having them available for alternative uses. Thus, economic costs can be incurred without any corresponding expenditure. For example, informal carers or volunteers may not cost anything, but they have an opportunity cost, because they could otherwise have used their time for leisure or employment. Some items may be free to service providers, but still have an opportunity cost (e.g. having rent-free use of a room or building, which could otherwise have been used by someone else).

There are four broad categories of cost that need to be considered as part of an economic evaluation:

- *Programme costs*: the direct costs of providing a service, which will include the cost of all the individual elements of

the service, such as staff costs, volunteer time, buildings, equipment, transport, support services, etc.

- *Non-programme costs (or savings)*: any resulting service effects, such as the savings that may result due to a reduction in the need for alternative services.
- *Patient and family costs*: any user or carer costs incurred as a result of consuming the services, such as travel to and from services, or child-care arrangements.
- *Productivity costs*: costs that result from an impaired ability to work as a result of illness or disability, lost economic productivity due to premature death or economic losses incurred by informal carers, who may have to reduce their work hours or give up work altogether as a result of their caregiving activities.

Most evaluation studies do not consider costs and those that do tend to focus on the more 'direct' programme costs. In part, this is due to evaluators not being fully aware of the economists' notion of costs and, therefore, neglecting costs that do not involve direct expenditure. But, there are a number of other reasons, which we discuss in turn.

### Narrow perspective

Our analysis of the economic literature in social welfare suggests that many evaluation studies adopt the perspective of the organisation funding or administering the programme. They may be interested only in the impact on their own budget. But, as noted earlier, economic evaluation should attempt to adopt a societal perspective, including costs incurred by all

those affected indirectly by the programme. Partial analyses can lead to misleading conclusions about the cost-effectiveness of certain programmes. For example, a policy of cutting personal assistance to older people in the community may appear to be more attractive than it really is if an evaluation does not take into account possible increases in the use of other public services or the additional burden borne by informal carers.

### Costing methodology

Some studies concentrate on measuring outcomes and examine costs only briefly, often tacked on to the end of an evaluation report. Yet, these studies sometimes make strong claims about the cost-effectiveness of the interventions they are evaluating. Costing is often incorrectly seen to be a simple exercise, whereas it can be very complex and few studies do it properly. Accuracy of cost data requires separate consideration of both the resources used and their unit cost. Regardless of the intervention being evaluated, the key steps involved are as follows: (Beecham, 2000):

- describing the ingredients of the programme, including the facilities, staff and other resources used as part of the programme
- identifying the different activities associated with the programme: the aim here is to list everything the service does and to decide on a suitable unit of measurement (e.g. hours of care or number of parent training sessions). Many social welfare interventions will involve a range of activities, which may need to be costed separately

- estimating the amount of resources used in each of these activities
- calculating unit costs (e.g. the cost per hour of care provided).

### Identification problems

Asking service providers to record clients' use of their services should be achievable, though even this level of information is often lacking. A balance needs to be found between what is desirable for evaluation purposes and what is practical 'on the ground'. Practitioners are often hard-pressed for time to carry out these activities and they may not like the feeling of being 'monitored'.

Identifying the non-programme costs and consequences of an intervention is likely to be even harder. First, it may be difficult to predict the nature of these impacts in advance, especially if these are unintended effects. Evaluators may need to carry out a scoping study to identify possible knock-on effects, for example by talking to practitioners and to clients. Second, some of these effects may be very indirect and, therefore, hard to attribute to the programme. For this reason, it is important to collect the same data for a control or comparison group to allow comparisons with the intervention group. Third, this information is unlikely to be collected as part of the administration of the programme, so a special survey may need to be carried out.

### Data problems

If conducted retrospectively, there is no opportunity to influence the type of data collected, so that certain basic cost data may be unavailable. The cost of data collection is

another possible constraint. Evaluators may rely on 'top-down' costing, where total costs are apportioned between specific activities or services, although this approach may underestimate costs (Beck *et al.*, 1999). Ideally, a more intensive 'bottom-up' analysis of all the resources used should be undertaken. For many social welfare evaluations, this would involve tracking users and recording both the services they come into contact with and the amount of time they spend with each service. A practical example of this approach is the Client Service Receipt Inventory (CSRI), which was originally designed for evaluations in health and social care, but can be adapted to other settings. The CSRI can also be used to identify the non-monetary costs of programmes, such as the use of informal care or the impact on people's ability to work (Beecham and Knapp, 1992).

With constrained budgets, evaluators may have to make judgements about the level of detail required. For example, a detailed cost analysis of local schemes may be unnecessary if the results are to inform a broad policy debate. It may be reasonable to make some general assumptions in place of collecting large amounts of cost data on specific schemes, which may not in any case be representative of other settings. In other cases, certain cost components can safely be omitted if these are relatively minor or if these costs are unlikely to differ across the programmes being compared. In addition, evaluators can make use of existing data and modelling to estimate and extrapolate costs.

### Valuation issues

We recommend a 'bottom-up' approach to estimating the unit cost of key activities (i.e.

those that account for a significant share of total programme expenditure or are an important focus of the evaluation). This involves collating costs for all relevant inputs to a service, such as staff time, buildings, equipment and management overheads. Some, though by no means all, providers may already have calculated the unit costs of their own services and national unit costs are available where more detailed information cannot be found. National unit costs are also recommended for costing services that add little to the total cost of a service (see, for example, Netten and Curtis, 2000 for unit costs of health and social care services).

Valuing some costs can be particularly problematic, such as the time costs of informal carers or voluntary workers. The difficulties involved help to explain why these costs are often omitted from evaluations. Progress has been made in valuing the economic costs of unpaid informal and volunteer activity, which are likely to be significant inputs into many social welfare interventions. The methodology now available, however, is not without its problems. For example, in practice, it can be difficult for individuals to quantify with accuracy the amount of time spent on caregiving activities. The appropriate unit cost for informal activity is not straightforward either. Often a single wage rate, such as the minimum wage or the cost of employing a domestic helper, is used as a proxy value for the opportunity costs of informal carers and volunteers. However, each individual will have a different opportunity cost of time, as they will have alternative uses for the time foregone. Few, if any, evaluations have incorporated this issue into the valuation of informal care, and

published studies are inconsistent in the methods used and the assumptions made (McDaid, 2001).

### Conclusions

Accurate estimation of costs is an essential building block in economic evaluation, yet few studies provide comprehensive information on costs, partly because they lack the incentive to

look at wider costs and partly because of some of the methodological issues we have discussed. The cost side of economic evaluation needs to be given more emphasis than is usually the case and needs to be treated as a more integral part of evaluation studies, rather than as an 'optional' add-on. Box 7 discusses the costing issues from two evaluation studies, including the foster care case study.

#### Box 7 Costing issues

##### *Foster care case study*

The foster care case study is a large study spanning seven local authorities and involving multiple agencies. The valuation of costs is complicated by the number of services involved and constrained by available resources (e.g. for data collection). Although fairly detailed data will be available on the use of social services from records, these will be limited to such information as the number of days spent in residential or foster care. The retrospective nature of data collection means that detailed information will not be available on the intensity of local authority placements and resources do not permit data on individual payments to carers to be collected. A detailed bottom-up costing for each of the 500 or so participants is, therefore, not possible and average local authority costs of care and accommodation will be used instead. This is an important limitation since the costs of foster and residential care vary greatly across placements and local authorities.

The level of detail on the use of other sector services is even more limited, because these are being collected using postal questionnaires. Respondents were simply asked to report the number of contacts with certain health, social and education services, as well as contacts with the criminal justice system. This will include, for example, the number of days spent in hospital for psychiatric or other reasons and the number of contacts with a social worker or educational psychologist. National average unit costs will be applied, because there is insufficient information to carry out more detailed costings.

##### *Home-based social work intervention for children and young people who have deliberately poisoned themselves*

This study of a social work intervention provides an example of a more detailed costing exercise within a complex area that spans many agencies. Information on the use of all health, education, social and voluntary services was collected from parents at the six-month follow-up

interview, using a questionnaire designed specifically for this purpose. In addition, an audit of medical records was carried out to verify the data on NHS contacts, in order to ensure greater accuracy of some of the key cost elements, such as hospital contacts.

The main focus of this study was on the social work intervention and the use of hospital services, so attempts were made to calculate accurate unit costs for these services. The study paid most attention to costing the social work intervention. Unit costs for 'assessment sessions' and 'intervention sessions' were calculated on the basis of a bottom-up estimation of the time the therapists spent on face-to-face and non-face-to-face contact, travel and supervision. An hourly rate was calculated from the midpoint of the relevant salary scales, plus employers' national insurance and superannuation contributions and an element added to cover capital and managerial overheads. Unit costs of hospital services were collected from local hospitals. We contacted each finance department in order to ensure that the costs provided were as detailed as possible.

For services that were less central to the study and those that added little to the total cost, less detailed methods of costing were used. For example, the unit cost of GP services was based on national estimates produced annually by the Personal Social Services Research Unit (see Netten and Curtis, 2000 for the most recent version).

(Source: Byford *et al.*, 1999)

## 8 Study design and methods

### Study design

Quantitative evaluators, including economists, generally favour the use of experimental approaches. These attempt to measure outcomes and, in the case of economic evaluation, costs for a group who have been 'exposed' to a particular intervention and a control or comparison group who have not. Comparisons between the two groups are used to isolate the impact of the intervention from other factors that may affect outcomes and costs over the evaluation period. The 'classic' experimental design is the randomised controlled trial, where patients are randomly assigned to receive the intervention being evaluated (e.g. a new drug) or an alternative treatment (e.g. a placebo). But, other study designs are more common outside the health care sector and are also considered here, including quasi-experimental studies and comparator groups. Where experimental approaches are infeasible or inadequate, it is sometimes possible to model the effects of a programme. Economists rarely use qualitative methods, but they are discussed here, because we believe they ought to play a more significant role in the economic evaluation of complex social welfare interventions.

Each of these approaches is discussed in more detail. But, we want to avoid presenting them as competing approaches, since they can (and should) often be used in combination with one another. Nor do we believe it is helpful in this context to think in terms of a hierarchy with randomised controlled trials (or RCTs) at the top followed by other less 'rigorous' approaches, where RCTs are not achievable. While this is the view widely held by health economists (e.g. NHS Centre for Reviews of Research and

Dissemination, 2001), we do not feel it translates to the social welfare field, for reasons that will hopefully become clear. In addition, we want to avoid concentrating overly on just one dimension of the study design. Often, too much attention is focused on the selection of a comparison group at the expense of other aspects of evaluation. It is possible to have a poor RCT-based evaluation, because insufficient thought has been given to the choice of outcomes or to how these are measured.

### Need for a comparator

It is important to have a comparison group, because it cannot usually be assumed that outcomes would remain the same in the absence of an intervention. Without a satisfactory comparison group, it is much harder to claim that changes in outcomes are due to the intervention and not to other 'external' factors, severely undermining the validity of a study's findings. This is especially important if there are good reasons to believe that changes in outcomes might occur 'naturally' over the course of time. For example, in evaluating a welfare-to-work programme, it is likely that some participants would have found employment without the help of the programme (Gardiner, 1997). Similarly, families are likely to enter mediation or other family-based services in a state of crisis, but will tend to regress towards less stressful functioning even without treatment (Pecora *et al.*, 1995). Without a comparator, it would be impossible to distinguish these natural changes from the impact of the intervention.

Before deciding how to recruit a comparison group, it is important to consider what the intervention being evaluated should be

compared with. The ideal comparison is with the next best alternative, which will usually be a 'do-nothing' option or current practice (i.e. an existing alternative), although this also depends on the specific evaluation question. For example, the evaluators of an alcohol rehabilitation centre wanted to assess whether shortening a residential programme would significantly reduce the effectiveness of the treatment, so it made sense to compare the shorter regime against the current regime, rather than a 'do-nothing' option (Long *et al.*, 1998).

In some cases, it may not be possible to have a clearly defined comparator. In our Safe In The City case study, for example, it is likely that most young people at risk of homelessness will be receiving some services. This would almost certainly be the case if the comparison group were recruited through social services, educational welfare officers, or other schemes such as Connexions, since these young people have already been picked up by the 'system'; those who slip through the net, and who would make a 'purer' comparison group, are unlikely to be picked up by the research team either. So, any comparisons that can be made will be between the SITC programme and 'current practice' in other areas. Current practice will itself vary enormously, depending on the circumstances of the young person and the local authority in which they live, so the comparator is a 'messy' one. Given that current practice may in some cases be quite similar to what SITC is doing, this will also make it harder to identify the additional impact of this programme. It would not be possible to assess whether the programme is effective in improving outcomes for young people, only whether it is

significantly more effective (or cost-effective) than current service provision elsewhere in London.

As well as a comparison group, it is possible to have more than one intervention group in order to test programme variants (e.g. more or less intensive versions of the same service or different combinations of services), although this does require a larger sample size.

### Randomised controlled trials

RCTs are commonly used to test the clinical effectiveness of pharmaceuticals and other health care interventions. Health economists have increasingly made use of them as the basis for economic evaluation, by adding cost data to new or existing RCT studies (Sassi, 2000). Their use is spreading to many other types of health care intervention and to other areas of social welfare (see Box 8 for some examples).

Randomised controlled trials (RCTs) are considered to be the 'gold standard' in health care evaluations, because well-conducted RCTs produce accurate (or unbiased) estimates of costs and outcomes. Differences in outcomes between the intervention and control group can be attributed with confidence to the intervention being evaluated, because any other differences between the two groups will be random and small, provided the sample is adequate. Reviews of evaluation studies in a number of social welfare areas have shown that studies with non-randomised control groups appear to systematically over-estimate the effectiveness of interventions (Oakley and Fullerton, 1996). For this reason, evidence from RCTs is usually rated more highly than evidence from other evaluation studies. In the health care field, reviews will often omit studies that are not

### Box 8 Examples of randomised experiments

#### *Example 1: evaluation of the New Deal for Lone Parents (NDLP)*

The NDLP is part of the Government's welfare-to-work programme aimed at encouraging more lone parents into work. It is a voluntary programme, whereby lone parents with school-age children are invited to an interview with a personal adviser and given information and advice on employment opportunities and financial assistance available to working mothers. A pilot scheme was launched in 1997 and evaluated before it was rolled out nationally. Although this is not strictly a randomised controlled trial, this study was set up to mimic one as far as possible.

The key issue for the evaluation team was how to identify the additional impact of the programme, given that many of the lone parents who found employment during this period would have done so even in the absence of this programme. Ideally, you would want to compare participants with non-participants, keeping all other factors constant. Fortunately, the design of the scheme created such an opportunity. The caseload was such that not all lone parents could be invited to participate in the scheme at once. Instead, invites were staggered over the pilot period, based on the last digit of an individual's National Insurance number, which is random. The researchers were able to estimate the impact of the programme by comparing changes in the employment status of those who were invited to join the scheme early on with an equivalent group of lone parents who were sent their invites at a later date (on average six months later), who acted as the control group. Data on over 15,000 cases were taken from administrative databases, which contain information on the claimant status of all individuals.

After six months, the probability that a lone parent would no longer be claiming Income Support was significantly greater for those who had been invited to participate in the programme than for those who had not yet been invited. Extrapolated over time, these results suggest that the overall stock of lone parents still on Income Support after one year would be reduced from 85 to 83 per cent as a result of the NDLP. Thus, the study was able to identify a relatively small, but significant, impact on the number of benefit claimants.

(Source: Hasluck *et al.*, 2000)

#### *Example 2: costs and benefits of community postnatal support workers*

This is an evaluation of a postnatal support scheme in Sheffield, based on a randomised controlled trial of 623 new mothers. If individuals agreed to participate in the programme, they were assigned randomly to a control group or the intervention group. The control group received the usual service offered by community midwives and health visitors, while the intervention group was, in addition, offered the services of a community support worker. The intervention consisted of up to ten visits within the first 28 days after the birth, providing



practical and emotional support to mothers. This support was tailored to individual mothers, but included housework, advice on breastfeeding and help with looking after other children.

The main outcome measure was the mother's general health perception after six weeks (using the SF-36 scale). In addition, a number of secondary outcome measures were used, including the Edinburgh Postnatal Depression Scale and breastfeeding rates. Use of health and social services was assessed over the same period, employing a client questionnaire. A follow-up survey of costs and outcomes was carried out after six months.

The results were analysed within a cost-consequences framework. The authors found no significant differences in the primary health outcome at six weeks and some evidence that secondary outcome measures were better for the control group. However, the women who received the service were overwhelmingly positive about it, more so than for the other maternity services they received. The total cost of the intervention was estimated at £180 per woman and the study found no significant effects on the use of other NHS services.

(Source: Morrell *et al.*, 2000b)

based on RCTs (e.g. NHS Centre for Reviews of Research and Dissemination, 2001). Example 1 in Box 8 is a good example of the advantages of a large-scale controlled experiment. Without randomisation it would not have been possible to identify a small, but significant, impact of the New Deal for Lone Parents, which was sufficient to justify this programme on cost-benefit grounds.

### *Practical and ethical issues*

There is resistance to using RCTs in many areas of social welfare, both among practitioners and evaluators who specialise in this field (Oakley and Fullerton, 1996). Most arguments against RCTs focus on practical or ethical issues, although these are not always clearly articulated:

- Some experiments are unethical (e.g. withholding potentially life-saving or life-changing services from some people) or may seem unethical (e.g. denying some people access to a well-established

intervention, even if its effectiveness is unproven).

- Sample size may be too small for randomisation to be effective (e.g. for very small-scale programmes or for area-based initiatives where whole areas, as opposed to individuals, would need to be randomly assigned).
- Evaluators may not have control over who receives a particular service and those who do may be reluctant to use random assignment, because they may have other priorities (e.g. to target their service at those who need it most).
- Randomised experiments may be 'contaminated' (e.g. if those in the control group find other ways of gaining access to similar services) or may be invalidated (e.g. if a high proportion refuse to take part or drop out of the experiment).

- Some experiments can be very expensive (or may be perceived as such).
- Politics may get in the way: those commissioning research may be more interested in ‘ammunition’ to help with their public relations than in rigorous evidence, which may not support their ‘cause’.

In some cases, these problems are genuine and effectively preclude this approach. But, some of these concerns are either mistaken or exaggerated and can, and have, been overcome in practice where the will has been there. Some of the above arguments can be turned on their head. For example, it is arguably more ethical to investigate an unproven intervention, which may mean denying some people access to a service for a period, than to continue practising untested interventions on a wider population (Williams, 1992). Other problems can be addressed through careful design (e.g. to minimise sample attrition) or by modifying the design. For example, random assignment can be made more acceptable if individuals in the control group are offered an alternative intervention, rather than a placebo (although this also makes it harder to show that the intervention is more effective.) Alternatively, it may be possible to approximate an RCT design without random assignment. Examples of ‘near-randomisation’ are discussed in the section on quasi-experimental studies.

### *Appropriateness of RCTs*

It is also important to consider whether RCTs are always appropriate, even if practical and ethical concerns can be addressed satisfactorily. First, there are some specific concerns about the impact of randomisation on the validity and

generalisability of the results. Random assignment may distort outcomes if people know which group they have been assigned to and if preferences play an important role in determining effectiveness. For example, those who have been assigned to the control group may be despondent, which may itself lead to worse outcomes, or they may be spurred on to do well against the odds, in which case outcomes for this group may be artificially inflated. This is a relatively easy hurdle to get over in drug trials (i.e. by using unlabelled bottles), but much harder in service evaluations.

Randomisation has also been shown to have an impact on who is willing to participate in the experiment. For example, participants in RCTs of preventative health care interventions tend to be more affluent, better educated and more likely to have adopted a healthy lifestyle than patients who decline (McKee *et al.*, 1999). The less representative the RCT sample, the less appropriate it is to generalise the results to the wider population.

Second, there are some more general concerns about experimental approaches to evaluation. We use the evaluation of a postnatal support intervention to illustrate some of these issues (see Example 2 in Box 8). This was a well-conducted RCT, published as a *Health Technology Assessment* report (Morrell *et al.*, 2000a) and in the *British Medical Journal* (Morrell *et al.*, 2000b), so our critique reflects the limitations of experimental approaches in this context, rather than the rigour of this particular study. Of course, it is always possible to argue that this study could have been designed differently and that these are not criticisms of experimental evaluation per se. However, we would argue that these are recurrent issues and that it is

naive to think that better design will always resolve these problems:

- Like other experimental studies, this evaluation relies on quantitative scales to measure outcomes such as the mother's general health and levels of depression. Any claims about this study need to be tempered by a note of caution about the appropriateness of these outcome measures. Health outcomes after six weeks showed no positive effects and some negative effects for those who received the support service as compared to those who did not, yet participants were overwhelmingly positive about the scheme. The outcome measures may be accurate, but it is possible that they were not picking up positive outcomes that were valued by mothers. Although this ambiguity might be resolved by further studies using different outcome measures, it demonstrates the possible limitations of using quantitative scales, even where these have been validated.
- Experimental studies focus on net outcomes, but it would have been helpful to know why this intervention did not appear to be working; this would help to assess whether the same intervention might be effective in other settings or with a more specific target group and how the scheme could be made more effective. The researchers suggest that the scheme might have failed because it did not build up people's own coping skills, but this theory is speculative. This type of evaluation could not provide any evidence as to why the scheme seemed to

be ineffective, yet this is crucial to the overall assessment of the scheme's potential.

- Like most other experimental studies, outcomes are measured at discrete points in time, in this case after six weeks and six months. The timing of surveys is often rather arbitrary, even though this can have a significant impact on the results. In this study, the first follow-up survey was undertaken shortly after the intervention ended. The authors suggest that mothers may have been experiencing a 'withdrawal' effect, which might explain the negative outcomes for the intervention group. But, again, this is speculative; experimental approaches to evaluation cannot easily address these types of question.
- An experimental design effectively treats the intervention as if it were a single homogeneous service, even though the service consists of a number of very different activities. So, this approach is unable to inform decisions about the appropriate mix or timing of activities.

For these reasons, we believe that the superiority of RCTs needs to be qualified in certain respects. Even where they are appropriate and feasible, they will often need to be carried out in combination with other research methods. In the above study, for example, qualitative research methods could have been used to identify those outcomes that were important to these mothers, which would have informed the choice of outcome measures. As it stands, this evaluation suggests that the support scheme was not cost-effective, but

using more appropriate outcome measures might have produced a different conclusion. In addition, qualitative research could have been used to explain the results of the RCT and to test some of the authors' own interpretations, which they admit are speculative – for example, whether mothers were experiencing a withdrawal effect at the end of the intervention. Last, but not least, qualitative research would have helped to identify those elements of the scheme that were valued most and by whom; this information could have been used to improve the scheme and to target it more

effectively. Qualitative research methods are discussed later in this chapter.

### Quasi-experimental approaches

These are similar to RCTs, except that users are not randomly assigned to a control group. Box 9 gives two examples of quasi-experimental studies, including our foster care case study. This approach may be more acceptable, because it avoids some of the real and perceived problems with random assignment. In some cases, this is the only feasible approach, because random assignment may be impractical.

### Box 9 Examples of quasi-experimental studies

#### *Example 1: evaluation of the Darlington Community Care Project*

The Darlington Project was one of many pilot projects initiated in the 1980s and funded under the Care in the Community Initiative. The scheme offered intensive and better-coordinated support in a community setting for older people who would otherwise have remained in long-term hospital care.

The researchers used a quasi-experimental design, because of the administrative aspects of the project (e.g. random assignment would probably have been politically unacceptable) and because they wanted to evaluate the project as a system as well as its effects on individuals. The comparison group was made up of patients from long-stay wards in an adjacent health district. They were shown to have very similar characteristics to the project's clients, although statistical adjustments were also made to control for any differences between the intervention and comparison groups. Over 200 older people participated in the study.

Participants were interviewed shortly before they were discharged from hospital (or, in the case of the comparison group, around the time they would have been discharged) and then six and 12 months later. Outcome measures included scales of subjective well-being (e.g. general satisfaction, morale, depression and loneliness) and various behavioural indicators (e.g. physical disability, apathy, communication difficulties and social disturbance), as well as some quality of care indicators (e.g. perceived need for improvements in the level of care). Death rates were also monitored for both groups. In addition, outcomes for carers were assessed (e.g. care tasks undertaken, distress).

The results showed significant improvements in all but one of the indicators of subjective well-being and in the quality of care indicators (though not, perhaps unsurprisingly, in most of the

behavioural indicators). These positive outcomes were not achieved at the expense of higher levels of distress for their carers. Although there was a significantly higher death rate in the intervention group after six months, this was not evident at 12 months.

Compared with hospital care, the Darlington model was also less costly (by around £30 per week, under conservative assumptions). Thus, the project appears to be a cost-effective alternative to long-stay hospital care, providing benefits to both patients and carers at a lower cost.

*Example 2: foster care case study*

Local authority placement for young people in care is unusual in that it is an area of social welfare where random assignment is clearly not feasible. Even if, from a practical perspective, all placement types were readily available to all young people in need, which they are not, it would be unethical to randomly assign young people to a ‘lottery’ of living conditions. It would, for example, be unreasonable to ask a young person to go back home when he or she had left home because of irreparable difficulties with the birth family. Nor is it possible to ensure that the placement type does not change over time. For example, a stable and long-term foster care placement cannot be achieved if either the young person or the foster family are unhappy with the arrangements.

Although the focus of this study is on outcomes for children in foster care, inevitably young people in the study will switch between placement types, allowing comparisons of alternative living situations. Comparing costs and outcomes of different placement types in this way is fraught with potential biases. The characteristics of young people in each placement type (reunification home, foster care, adoption, residential care and frequent movers) are likely to vary significantly. Those who are adopted, for example, may be younger with fewer behavioural problems than the rest of the sample; those who return home may have had fewer problems to start with; and frequent movers may be the most disturbed.

Although a straightforward comparison between groups would be inappropriate, differences between groups can be explored using statistical (multivariate) analysis that helps to control for differences between groups. This is made possible by the large quantity of data on the characteristics of these young people, their placements and carers. In addition, qualitative data will be drawn on to help interpret and disentangle the results of the quantitative analysis.

Where random assignment is not possible, evaluators can take advantage of ‘natural’ variations in the implementation of programmes to locate a suitable comparison group, for example by comparing a school that

has implemented a new initiative with one that has not. If the circumstances are right, it may be possible to achieve something close to random assignment (see Box 10 for examples of ‘near-randomisation’).

### Box 10 Near-random assignment

#### *Waiting list control group*

A control group can sometimes be constructed from those on a waiting list for a programme. For example, in an evaluation of a parenting initiative, parents were assigned to the training class nearest to where they lived and, once each class was full, parents were placed in the waiting list control (Thompson *et al.*, 1996). This is a good approach provided the programme is over-subscribed and provided access to the scheme is on a first come, first served basis (i.e. places are allocated on a fairly random basis). One problem (for evaluators) is that those on the waiting list may join the programme before the end of the evaluation period. Therefore, this approach is more appropriate for programmes with fairly short-term outcomes or long waiting lists.

#### *Staggered implementation*

This is where the delivery of a programme is staggered over time, allowing comparisons to be made between those who receive the service sooner or later. An example is the evaluation of the New Deal for Lone Parents (see Box 8). Although invitations were sent out in a random order (based on the individuals' National Insurance number), lone parents who had not yet received an invitation could ask to join the scheme on their own initiative, so early participation in the scheme was not completely random. (This will lead to a slight under-estimate of the programme's impact, since some lone parents in the control group will also have benefited from the programme.)

(Source: Hasluck *et al.*, 2000)

#### *Natural variations*

In some situations, evaluators can take advantage of unplanned or 'natural' variations in the implementation of a programme that may be fairly random. An example of this is an evaluation of the Nokomis programme for young offenders (Deschenes and Greenwood, 1998), where the comparison group was made up of young offenders who continued to be sent to residential institutions, because some judges refused to make referrals to the new programme. Another example is an evaluation of the One-to-one early reading programme (Farkas, 1998). Some children attended fewer sessions because of scheduling problems, which were apparently unrelated to the characteristics of the children (i.e. they were fairly random). The impact of the programme was estimated by comparing outcomes for those who had attended only a small number of sessions (who acted as the comparison group) with those who completed the whole course.

#### *Cohort comparisons*

Another approach is to compare outcomes for different cohorts of service users before and after a change in the way a programme is delivered. This approach was used to evaluate the impact of a modified alcohol treatment programme. Outcomes for a group of clients who attended a particular alcohol treatment clinic were compared with outcomes for a second group of clients

who attended the same clinic after the programme was substantially shortened from five weeks to two weeks (Long *et al.*, 1998). For some chronic problems, it may be possible to compare outcomes for the same individuals following changes in their treatment over time.

### *Pragmatic approaches*

More pragmatic study designs are likely to involve some compromise on the 'ideal'. For example, those administering the programme may want to give priority to certain clients and, therefore, override the assignment process in some cases. An example of this is the evaluation of a parenting initiative (Thompson *et al.*, 1996). Although they had a waiting list control group (see Box 10), parents who felt they needed immediate help were allowed to join the scheme sooner. In some cases, these kinds of compromise may be forced on the evaluator. The national evaluation of Breakfast Clubs was based on random allocation of funding for the first wave of the programme, using those schools who did not receive funding as a control group. But, some of these schools received funding under the second wave and will have started their own Breakfast Clubs before the end of the evaluation period (Mugford, 2001). Instead of comparing schools with and without Breakfast Clubs, the study will now compare schools that introduced them earlier or later.

### *Recruiting a comparison group*

In some cases, it will be relatively straightforward to identify a potential comparison group, who are similar to those receiving the intervention. But, this is not always the case (see Box 11). Recruiting a comparison group is harder if:

- the target group is not defined in very specific terms (e.g. young people at risk of

homelessness), especially if there is discretion in the way the eligibility criteria are applied in practice

- the characteristics of the target group are not readily observable (e.g. people with mental health problems or children with reading problems), in which case some sort of screening process may be necessary
- the service is universal or very widely used, in which case it may be difficult to find a group of people who are not already receiving it (e.g. the new leaving care legislation)
- the target group is very small and/or widely dispersed, in which case it is more difficult (and more expensive) to locate sufficient people with similar characteristics
- there are no obvious referral points or sites where a comparison group can be recruited from or there are multiple referral points, which makes the recruitment process more complicated
- entry to the programme is based on self-selection (i.e. volunteers) or is carefully controlled by service providers, which introduces other confounding factors, such as the motivation of clients, which are hard to match but which can have a significant impact on outcomes.

### Box 11 Problems in recruiting a comparison group

#### *Safe In The City case study*

This was a particular problem for this case study and the same sorts of issues are likely to be an obstacle to experimental approaches in other areas of social welfare. Projects were very resistant to the possibility that some young people who had been referred to them could then be allocated (randomly or otherwise) to a comparison group and receive no services. This might be less of an issue if there were long waiting lists for these projects, but at this early stage in the programme's development, this was difficult to predict. When we visited them, some of the projects were relatively new or had recently expanded their service, so they were doing their best to develop good relationships with potential referral agencies – and they were concerned that referral agencies might be deterred if they knew some of their clients would be allocated to a comparison group. Another option was to construct a comparison group from those who had refused a place on the SITC programme or who had dropped out at an early stage. However, these young people are likely to be less motivated than those who complete the programme, so we could not assume that what happens to them is representative of what would have happened to their clients, had they not taken part in the programme.

Thus, the only viable option would be to recruit a comparison group from other London boroughs not covered by this programme. This would require some sort of 'mock' referrals system in other areas, whereby local service providers (e.g. Education Welfare Officers or Pupil Referral Units) would be asked to identify young people with similar characteristics to those who take part in the SITC programme and to pass on their details to the research team. The previous referrals system would be very difficult to replicate, because the criteria used for assessing possible clients were not clearly defined and varied between projects. However, the new, more standardised assessment process should make this easier in the future. But, this approach would still be very dependent on the cooperation of other agencies in these areas, who may not see the research as a priority.

#### *Potential bias*

Non-random control groups introduce possible bias. If the intervention and comparison groups are not identical, then at least some of the differences in outcomes might be accounted for by differences in the characteristics of the two groups, rather than the intervention itself. This is not just a theoretical possibility. Where participation is voluntary or where clients are selected into a programme, there are good

reasons why the groups may differ in ways that are difficult to observe. For example, service providers that are being judged on the outcomes achieved by their clients may selectively admit those who, in their experience, are most likely to do well.

It is important, therefore, to identify and, where possible, control for any differences between the characteristics of the intervention and comparison groups. More sophisticated



quasi-experiments seek to minimise or control for potential bias in various ways (see Box 12). These techniques are more likely to be successful if information is collected on as many

relevant variables as possible (i.e. those that may have a significant impact on outcomes). Sample size will affect how many of these factors can be controlled for in practice.

### Box 12 Adjusting for potential bias

#### Example 1: matching controls

When the comparison group is non-random, matching can help to minimise differences between the intervention and comparison group. A good example is an evaluation of an alcohol treatment programme in the US, which compared two groups or cohorts of clients before and after a significant shortening of the treatment programme. A straight comparison between the two groups might have been misleading. For example, if the intervention group contained a higher proportion of clients with less severe alcohol dependency, this would have biased the results in favour of the intervention group. Instead, the researchers tried to match each patient in the intervention group (i.e. the shorter programme) with someone in the comparison group with similar characteristics, on the basis of gender, age and severity of dependence on alcohol. This process resulted in 75 pairs (i.e. a total of 150 observations out of an original sample of 212). A comparison between the matched groups showed that abstinence or non-problem drinking was achieved by just over half of all clients in each group (after one year). Thus, the change in programme delivery did not appear to affect outcomes, but treatment costs were significantly reduced. Hence, the authors concluded that the new treatment programme was more cost-effective.

(Source: Long *et al.*, 1998)

#### Example 2: statistical methods

This evaluation of two intensive community-based interventions for juvenile offenders is an example of the use of statistical methods to control for potential bias. The study examined the impact of each intervention on criminal justice expenditures and compared this with regular probation (i.e. whether more intensive interventions were successful in reducing future offending and the associated costs to the criminal justice system). Three youth courts in Mississippi each undertook one of the three interventions, so random assignment of participants was not possible. In addition, the dropout rate was very high (just under 50 per cent), which introduced another potential source of bias.

Statistical analysis was used to strengthen this quasi-experimental design. First, the researchers estimated a regression equation to predict the probability of someone completing the programme, given their characteristics (e.g. age, previous contacts with the criminal justice system and various behavioural indicators). This adjusts for any bias in the results caused by certain types of client dropping out of the programme and out of the evaluation study.

*continued overleaf*

The researchers then estimated a second equation to predict criminal justice expenditures as a function of the client's characteristics and the type of intervention they received (for those who completed the programme). If spending depends only on the clients' characteristics, then the variable denoting the intervention they received would be insignificant. However, the results indicated that one of the two intensive interventions (cognitive behavioural therapy) did significantly reduce criminal justice expenditures during the 18-month period, after adjusting for differences in the characteristics of young people in each group. The estimated savings were just under \$3,000 per participant, which is around double the estimated additional cost of this intervention. As well as controlling for potential bias, this type of analysis strengthens the plausibility of the study's results, because coefficients on the key variables in the model were consistent with prior expectations (e.g. higher spending on those who have had previous contacts with the criminal justice system).

(Source: Robertson *et al.*, 2001)

Quasi-experiments are usually seen to be a second best to RCTs, but it is not often discussed whether second best is good enough. This depends on how far potential biases can be controlled for within a quasi-experimental study. In the welfare-to-work field, quasi-experimental designs are sometimes rejected, because they cannot control for the motivation or attitudes of clients, which is an unquantifiable, but potentially important, source of bias (Payne, 2000). It is less problematic if key characteristics are observable and measurable or if the potential bias is small relative to expected impact of programme. Unfortunately, the results of studies are frequently presented without explicitly considering potential biases (see section on 'Interpretation and presentation of results' in Chapter 9).

### Comparator groups

Quasi-experimental studies should involve some attempt to control for differences in the characteristics of the intervention and

comparison group. Where this is not possible, then it is perhaps more appropriate to talk in terms of comparator groups to distinguish them from matched or statistically controlled comparison groups. Comparator groups are often used with programmes that are directed at whole areas or institutions, when sample size is too small to match the groups accurately or to carry out statistical analysis. Box 13 provides an example of a study of school-based social work, using a comparator group.

This may be a deliberate choice to restrict the number of areas or institutions, so that these can be studied in more depth. This needs to be weighed up against the advantages of including a larger, and therefore more representative, sample. For more complex interventions, in particular, it is important not only to look at measurable outcomes, but also to understand the way the programme is operating and why it may or may not be effective. This is more like a case study analysis, using qualitative information to help interpret quantitative data. A good example of this is an evaluation of the

Priority Estates Project (PEP) (Glennister and Turner, 1993). A straight comparison of the Hull experimental estate and the comparator estate showed few differences following the introduction of devolved housing management. But, this masked significant differences *within* the Hull estate. One part of the estate was doing very well, at least in part because of beneficial changes brought about by the PEP scheme.

Meanwhile, one particular block on the estate had deteriorated markedly because a large number of young and formerly homeless people had been allocated there. Thus, any beneficial effects of PEP were being offset by a change in the local authority's allocations policy – an external influence that is unlikely to have been picked up in a more superficial analysis of a larger number of estates.

### Box 13 Example of an economic evaluation using a comparator group

#### *An experimental study of school social work with cost-benefit analyses*

This was a three-year experimental programme of school social work involving one part-time and two full-time social workers in a primary school and secondary school serving deprived council estates in Dorset, England.

The two project schools were compared with another primary and secondary school in the same area. Data were collected on 1,300 children and adolescents, including truancy, self-reported theft, bullying, drug use, vandalism and net exclusions (i.e. numbers of pupils excluded from a school less the number of excluded pupils transferred into the school). In addition, qualitative interviews were carried out with all the teachers at the project schools and a small sample of parents and young people.

The results reveal significant reductions in six out of the seven indicators examined, including net exclusions. Based on the estimated cost savings from a reduction in net exclusions, the authors conclude that the experiment was 'highly cost-effective'. Their qualitative analysis appears to support these findings, suggesting that school social work can be effective in improving the school climate and teacher morale.

Though these results are clearly promising, there are some concerns about this study, which the authors acknowledge. It became apparent during the course of the evaluation that one of the comparator schools was probably not the best school to choose and that in hindsight (and with the availability of new demographic and social services data) a different comparator should have been chosen. Thus, other differences between these schools may account for the better performance of the project schools over this period. Further evaluations are needed to confirm these initial findings in other schools, with different social workers and with more appropriate comparator schools.

(Source: Bagley and Pritchard, 1998)

It may be difficult to find a good comparator because no two areas or institutions are the same – and there are many different factors to take into account. Even if two areas or institutions are very similar at the beginning of the evaluation period, circumstances may change during the evaluation period. Alternatively, the ‘experiment’ may be contaminated by changes in policy within the areas or institutions being studied. For example, in the evaluation of the Priority Estates Project, those managing one of the comparator estates decided to introduce some elements of the PEP scheme during the course of this evaluation.

There is much more scope for bias with comparator groups, which are non-random and generally small-scale, so it is even more important that alternative explanations for the results are considered. While this approach is clearly better than having no comparison group, this kind of study needs to be supported by other evidence to link the intervention to observed outcomes. It cannot be assumed that any differences between the intervention group and the comparator(s) are necessarily due to the intervention.

Small-scale studies add to the body of knowledge on a particular intervention. Over time, it should be possible to build a more accurate and complete picture, if, for example, study after study produces similar or consistent results. It may even be possible to carry out a meta-analysis if there are sufficient studies (see later in this chapter).

### **No control or comparator group**

Many evaluation studies involve a before and after assessment of outcomes, but without a comparison group. Indeed, these types of study

dominate the effectiveness literature in areas such as social work (MacDonald, 1996). Studies of this kind rarely give results that can be used as part of an economic evaluation, unless there is strong evidence that there would have been no change in the absence of the intervention. The results need to be interpreted with great care; at best, they are suggestive. Confidence in the results can be improved if several such studies produce similar results for a range of users and in a range of different circumstances (MacDonald, 1996). But, these types of study should be used in an exploratory way, for example to provide justification for a more rigorous (and expensive) evaluation of an intervention, rather than as definitive evidence of an intervention’s effectiveness or cost-effectiveness.

### **Complementary research methods**

Other research methods, where they are employed as part of an economic evaluation, will most likely be used to complement one of the experimental approaches discussed in the previous section. In general, the more complex the intervention, the more likely it is that a pure experimental evaluation will be insufficient on its own. This was certainly true for our case studies, all of which incorporated other research methods.

### **Modelling**

Models make use of secondary data and other sources, including expert opinion, to predict the outcomes or costs of a particular intervention within a well-structured quantitative framework. All models involve a set of inputs and assumptions and a more or less complex set of calculations for generating predicted outcomes or

costs. For example, large-scale databases with information on individuals' medical records and cholesterol levels have been used to model the long-term benefits of interventions designed to reduce patients' cholesterol levels (e.g. Stinnett *et al.*, 1996). In the social welfare field, modelling is used extensively to estimate the impact of changes in the tax-benefit system (Mitton *et al.*, 2000). One of our case studies uses a model to estimate how different households might

respond to changes in the energy efficiency of their homes (see Box 14).

Modelling can be used in a range of situations on the back of, or in place of, other approaches (Buxton *et al.*, 1997):

- as part of a scoping study to help determine which variables are likely to be of greatest importance and should be given most attention in designing an evaluation study

#### Box 14 Example of the use of modelling

##### *Fuel poverty case study*

One of the objectives of the Home Energy Efficiency Scheme is to reduce the number of people living in cold homes. Yet, relatively little is known about how people adjust the warmth of their home in response to energy efficiency improvements. Ideally, information would be collected on the temperature in people's homes before and after improvements brought about by the HEES scheme. But, in the absence of this kind of data, we attempted to estimate the likely response, using data from a representative survey of UK households, containing information on 7,000 households (including temperature readings and energy efficiency ratings for each home).

Our model was based on simple economic theory. Economic analysis of consumer decisions shows that an individual's demand for a product depends on its price, and on the individual's income, personal tastes and other characteristics. Other things being equal, demand for a good or service will increase if its price falls or income rises. This basic structure can be applied to decisions about the temperatures people set in their homes – their 'demand for warmth'. People decide how warm they would like their home based on how much it costs to heat it, what their income is and other factors (e.g. if they have young children, they may choose a higher temperature). We were particularly interested in how sensitive people's demand for warmth is to changes in the energy efficiency of their homes, which is equivalent to the 'price of warmth'. By lowering heating costs, energy efficiency improvements should increase the temperature in people's homes, but by how much?

We estimated the demand for warmth equation using statistical (regression) analysis on data from the English House Condition Survey (for 1991 and 1996). The dependent variable, which is what we wanted to explain, was the temperature in the living room at the time the survey was carried out. The key explanatory variable was the 'price of warmth'; this is the estimated

*continued overleaf*

daily cost of heating the dwelling to the standard regime (18°C in the hall and 21°C in the living room) and is directly related to the energy efficiency of the property. Other variables that might influence people's demand for warmth were added to the regression equation, including household income, tenure and presence of an elderly person or infant. We ran a long list of regressions, but the one that performed best was a simple linear regression between living room temperatures and standardised heating costs, with dummy variables for each income group and for other variables that were found to be significant. Extensive checks were carried out to test the robustness of the results, including trying out non-linear forms and excluding certain observations (e.g. where survey respondents had said the temperature was different to normal).

The key result is that we found a significant and negative 'price' effect, as expected. The size of this coefficient implies that a £1 reduction in the daily heating costs is associated with 0.35°C increase in the living room temperature. This is a relatively small effect, but this is what we would expect for a necessity like warmth. Nor is there evidence that poorer or elderly households, who are the main target groups for this scheme, are likely to be any more (or less) responsive than other types of household. The estimates from this model could be used to simulate the impact of HEES on the number of cold homes in England.

- to extrapolate beyond data on intermediate outcomes in order to estimate final outcomes (e.g. life expectancy based on reduced short-term morbidity) or to extrapolate the same outcome beyond the time period covered by an evaluation (e.g. lifetime criminal behaviour, based on juvenile crime records)
- to examine variations in costs and outcomes between population sub-groups (see section on 'Analysis of results' in Chapter 9)
- instead of an experimental evaluation, where this is infeasible; modelling is particularly useful for evaluating programmes retrospectively or for simulating the impact of a new

programme (or changes to an existing programme) in advance.

The advantage of models is that they are flexible and cheap. They are able to combine information from a wide range of sources and can be used to test the sensitivity of the results to variations in the underlying assumptions. Models do not rely on data being collected directly from users, but make use of existing data sources, including administrative databases, large-scale household surveys and results of previously published evaluations. Where there are information gaps, assumptions can be made until further evidence becomes available.

However, the results from modelling are only as good as the quality of the data and the validity of the underlying assumptions. Models are not a substitute for lack of data, although they help to

make the best use of limited data. In many areas of social welfare, the scope for modelling will be constrained by lack of data. Another disadvantage of models is that they may not be transparent to those who were not involved in designing them, including decision-makers, who may not be fully aware of the assumptions on which they are based (Sassi, 2000).

### Qualitative research methods

Qualitative research covers a wide range of methods (see Box 15). Unlike the methods discussed above, the aim is not to provide quantified answers, but to focus on the experiences, views, meanings and motivations of individuals and organisations. As such, qualitative research is useful for addressing a different set of evaluation questions: not ‘how many  $x$ s?’, but ‘what is  $x$ , how does  $x$  vary in different situations and why?’ (Pope and Mays, 1995). There are a number of ways in which qualitative methods can be used as part of an economic evaluation (e.g. Coast, 1999):

- as a preliminary step to help determine the scope and design of a quantitative study; for example, they can be used to identify outcomes that are seen to be important by practitioners or users (and that should, therefore, be included in the evaluation) and ways of measuring these outcomes (e.g. helping to develop survey instruments)
- to provide a qualitative assessment of complex outcomes that are difficult to quantify (e.g. changes in organisational culture or sense of community), which can be presented alongside quantified outcomes

- to help interpret the results of a quantitative study (e.g. why certain outcomes are observed, in particular unexpected or apparently inconsistent results)
- to assess whether an intervention seems to be working as intended and, taking this a step further, to provide qualitative evidence to link the intervention to observed outcomes
- to gain a better understanding of why an intervention is working (or not) in a particular setting and, therefore, some insight into whether, and under what circumstances, the results of a quantitative study might hold more widely.

At present, qualitative research methods are rarely used by economists. One reason, other than lack of familiarity, is a concern about the scientific rigour of these methods. But, this results largely from a misunderstanding about the purpose of qualitative research (Mays and Pope, 1995a). The aim is not to generate statistically significant results, but to explore themes, patterns and associations within a richer and more diverse data set. Where a quantitative researcher would attempt to show causality by looking for statistical correlations between variables, a qualitative researcher would look for explanations – those given by the interviewees themselves, but also those inferred by the researchers on the basis of people’s accounts of what happened and how they responded.

There is a danger of economists using these approaches naively (Coast, 1999). But, used

### Box 15 Qualitative research methods

#### *In-depth interviews*

These are face-to-face conversations with the purpose of exploring issues or topics in depth. They may be based on a loose structure of open-ended questions that define the area to be explored, at least initially, and from which the interviewer or interviewee can diverge in order to pursue an idea in more detail. Qualitative interviews are interactive and flexible. They aim to get below the surface of the topic being discussed, explore what people say in more detail and from the interviewee's perspective, and uncover new areas or ideas that were not anticipated at the outset. They are particularly useful for exploring people's experiences or behaviour, their opinions or beliefs, their feelings, or their knowledge on a subject.

(Source: Britten, 1995)

#### *Focus groups*

These are a method of group interview that explicitly includes and uses group interaction. Instead of asking each person to respond to a question in turn, the group are encouraged to talk to one another: asking questions, exchanging anecdotes and commenting on each other's experiences and points of view. The method is particularly useful for exploring people's experiences (e.g. of homelessness) or knowledge (e.g. of services available for homeless people) and can be used to examine not only what people think but also why they think that way. Group discussion is particularly appropriate when the interviewer wishes to encourage participants to explore issues of importance to them and using their own vocabulary.

(Source: Kitzinger, 1995)

#### *Observational studies*

These involve the researcher systematically watching people and events to find out about behaviours and interactions in natural settings (e.g. the point of service delivery). Researchers sometimes participate in the activity they are observing or, in some cases, may carry out their research covertly. An important advantage of observational studies is that they help to overcome discrepancies between what people say and what they actually do; they may even uncover behaviours or routines of which the participants themselves are unaware. Observational methods are particularly well suited to studying the way organisations work.

(Source: Mays and Pope, 1995b)

#### *Case studies*

These are in-depth studies of one or a small number of settings, where broad, intricate issues have to be addressed in a complex environment. Case studies typically use a range of different research methods, including both quantitative and qualitative approaches. Those that are largely qualitative are most valuable when the focus is on why and how the intervention



succeeds or fails, where the general context is likely to influence the outcome and where researchers have no control over events. Case study sites may be chosen to represent a range of different situations or to test a particular theory.

(Source: Keen and Packwood, 1995)

appropriately, qualitative research can help to address many of the shortcomings of experimental approaches. We have already given some examples of how qualitative research could have helped in the evaluation of a postnatal support scheme. Box 16 gives some more examples of their potential application, based on our case study evaluations. One of the challenges for economists working in this field is to find better ways to combine the results of quantitative and qualitative analysis within a common framework.

### Meta-analysis

Meta-analyses combine data from previous evaluations of the same intervention. They differ from systematic reviews in that they involve statistical analysis of data from earlier studies, as opposed to a more qualitative assessment of the evidence. The advantage of this approach is that it increases the validity of cost and outcome estimates, by increasing sample size, and can be used to explore variations in a programme's effectiveness (and cost-effectiveness) across a wider range of settings. A meta-analysis is also cheaper and quicker than conducting further evaluation studies. However, meta-analyses are feasible only if sufficient well-conducted studies of the intervention have already been carried out. In addition, meta-analyses tend to focus on published studies, which may generate a bias in favour of significant results. There is also a risk

of mixing 'good' and 'bad' studies, unless weaker studies are filtered out in some way.

Box 17 provides an example of a meta-analysis in the social welfare field. But, the use of meta-analysis is likely to be fairly limited in most areas of social welfare, at least in the foreseeable future, because there are relatively few rigorous evaluation studies on which to base them. Few programmes remain stable for long enough to build up a strong evidence base. Most meta-analyses would be reliant on US studies (see Chapter 3) and these will have less relevance for the UK, because of differences in the way social welfare services operate in the two countries.

### Other design issues

So much attention is given to the nature of the comparison group that other important design issues are sometimes neglected. Some of these are discussed in this section.

### Sample size

With any quantitative study, sample size is crucial. Many studies are underpowered, which means that the sample is not large enough to identify significant effects on outcomes and costs. Sample size is less of a problem if a programme's impact is expected to be very large. But, if the evaluation is trying to identify a small, but important, effect, then a large sample is required. If, for example, a study were

examining the impact of an intervention on a rare but serious condition (such as homelessness or violent crime), then there would be little point in carrying out a small-scale quantitative survey. The sample also needs to be larger if there is a lot of variation in costs and outcomes between individual users. Cost distributions are

typically skewed, with a few individuals accounting for a high proportion of service use, so economic evaluations may require a much larger sample in order to detect cost differences than studies that examine only outcomes (Briggs, 2000).

### Box 16 Examples of use of qualitative methods

#### *Fuel poverty case study*

While the main component of this case study was a model to estimate the impact of energy efficiency improvements on the temperatures in people's homes (see Box 14), we also used qualitative research methods to help interpret and qualify the results of our quantitative analysis. In-depth interviews were conducted with 18 households who had applied for the Home Energy Efficiency Scheme and were either waiting for work to be done or had recently had work done.

These supported the quantitative finding that people's demand for warmth is unlikely to be very responsive to changes in the energy efficiency of their home. Few, if any, households were willing to sacrifice comfort in order to save money. And, this was reflected in their behaviour: 'If it's cold, I'll put the heating on ... I'm not going to sit here and freeze'. But, it was also possible to gain a more detailed understanding of how people make decisions about their heating, and in particular how these are influenced by financial and other considerations, some of which were not picked up in our quantitative analysis.

Our qualitative data were used to develop a typology of household behaviours. For example, one group of households who live in cold homes see heating as an absolute priority and will do whatever they can to keep warm (e.g. using portable fan heaters, turning the oven on and leaving the door open). This group spend above-average amounts on their heating, but are fighting a losing battle to keep warm, because their heating system is inadequate. They stand to gain most from energy efficiency improvements, both in terms of substantial savings and improved comfort, which they are likely to value more highly than other households. Another group of households also live in inefficient homes, but use their system much more sparingly, perhaps heating just the one room and only when they 'need' it. They are prepared to make some compromise on their comfort, but financial considerations do not seem to be paramount. Some of these households may be so used to living in a cold home that they do not have a strong unmet demand for warmth. Although these households have the greatest scope to gain from energy efficiency measures, they may not take advantage of this, preferring to enjoy modest improvements in

comfort at little, or no, additional expense. Thus, qualitative analysis is helpful in identifying various ‘unquantifiable’ factors that may have a significant influence on how people respond to the HEES scheme, but that are omitted from the quantitative analysis.

Our qualitative analysis also suggested that our main outcome measure (i.e. a spot measure of the living room temperature) might not be a very good proxy for the warmth of people’s homes. The concerns people had about their heating were much more specific than this, including certain rooms or parts of the house that were too cold (almost never the living room), being cold at certain times of day and uneven temperatures (e.g. one woman talked about being ‘cold around the edges’). Thus, more sophisticated measures of warmth (or, better still, comfort) are needed. Because we used a narrow measure of warmth, our model is likely to underestimate the difference in comfort levels between ‘cold’ and ‘warm’ homes and may, therefore, underestimate the benefits of improving conditions in the worst homes.

### *Gatsby case study*

As part of our pilot study, we held a half-day discussion with each of the groups we assessed, structured around those aspects we had identified as being important to the effectiveness of community groups (i.e. clarity of vision, how well they worked as a team, how they were perceived by their local community and whether they had adequate support from external bodies). We used this to make our own assessment of groups against these criteria and to interpret the groups’ responses to our questionnaire (see Box 6). Thus, qualitative research methods were employed to inform our assessment of the groups’ effectiveness in the absence of more objective measures. We also used these visits to validate the criteria we were employing to assess groups (including asking groups what they thought made a good community group). Our criteria stood up well to this test and seemed to cover all the key areas of the groups’ functioning.

### *An experimental study of school social work*

A major part of this evaluation (see Box 13) was qualitative, including interviews with teachers at the project school and a sample of parents and children. This supported the findings from the quantitative analysis and provided evidence to link the intervention to the improvement in quantified outcomes, including a lot of positive feedback from young people who had been helped by the project’s social workers. These interviews also highlighted a number of themes that were associated with the perceived success of the project, including: young people feeling individualised, the importance of having informed advice available, confidentiality, trust and a calm approach. Another factor identified by the researchers was the ability of social workers to intervene immediately a problem arose. All these ‘lessons’ are useful in understanding why the intervention appears to have been a success and, therefore, which specific characteristics should be emulated by other school-based projects that are seeking to replicate this elsewhere.

### Box 17 Example of meta-analysis

#### *Are private prisons more cost-effective than public prisons?*

This is a meta-analysis of 33 cost-effectiveness studies of public and private prisons in the US, focusing on adult male prisons. It tests the theory that private prisons are more efficient because of competitive forces and fewer bureaucratic constraints. However, the results show that private prisons are no more cost-effective than public prisons, once other institutional characteristics are controlled for. This study demonstrates the benefits of carrying out a meta-analysis, especially where previous studies have produced mixed results. Previous studies had tended to focus on a small number of prisons – typically one private and one public facility – which made it hard to make broader inferences from the results of individual studies. Some studies also lacked methodological rigour; for example, they did not always consider other factors that might account for differences in the cost per inmate.

This meta-analysis was able to combine the results of independent studies to address a common research question. Estimates of cost per inmate for different prisons were regressed against an ownership variable (i.e. whether public or private) and a small number of other variables, including the age of the facility, the number of inmates and the security level. The ownership coefficient in the equation was not significant, implying that there are no systematic differences in the efficiency (i.e. cost per inmate) of public and private prisons. Institutional characteristics, including the size of the prison, its age and security level, were all found to be significant factors in explaining variations in cost per inmate.

(Source: Pratt and Maahs, 1999)

The implications of inadequate sample size depend on the nature of the interventions being compared. If the programme being evaluated is intended to produce better outcomes than existing interventions, then underpowered studies may fail to identify positive outcomes and so are less likely to support innovative or more intensive interventions. If, on the other hand, an existing programme is being compared with a cheaper, less intensive programme, then underpowered studies are less likely to identify a significant deterioration in outcomes. The danger is that studies may conclude that less intensive, cost-cutting, interventions are cost-

effective, but only because the sample size was too small to identify an adverse impact on outcomes.

One of the problems in designing an evaluation is that it is often hard to know in advance how large the sample should be, unless previous studies have already been carried out on similar interventions. As already noted, information is needed on the variability of costs and outcomes, but this information may only become available after an evaluation of the programme has been carried out. Sample size may also be constrained by data collection costs or by the size of the programme.

### Drop-out rates

Even if the original sample is large enough, a study can be undermined by people dropping out during the course of the evaluation. Some people may drop out of the programme that is being evaluated, but continue to be monitored as part of the evaluation study. This is less of a problem from an evaluation point of view. Unless there are good reasons for not doing so, these people should be included in the intervention group when the results are analysed. Excluding those who drop out would bias the results, because those who are able and motivated enough to complete the programme are probably more likely to do well. This is certainly true of welfare-to-work programmes and other interventions where the success of the programme is very dependent on the efforts of those taking part.

The downside of including those who have dropped out is that this will 'dilute' the impact of the programme, because some of those in the intervention group will not have received the full intervention. This seems justified if helping and encouraging clients to complete the programme is considered to be part of a successful intervention. However, people may drop out for other reasons that are unrelated to the effectiveness of the programme. For example, in an evaluation of the Revolving Doors Agency, a mental health intervention for people who have been in contact with the police (Revolving Doors Agency, 2000), some potential clients did not join the scheme because they were released outside normal working hours, not because project workers failed to engage them. In this case, it may be reasonable to exclude them from the analysis.

A different problem is where people drop out of the evaluation study. This will often be the same people who drop out of the programme, though not necessarily so. Evaluators often want to collect cost and outcome data after people have completed the programme, often long after, when it is generally much harder to locate participants. The obvious contact point (i.e. the point of service delivery) is no longer available and some people will have moved. This is a particular problem for social welfare programmes that are targeted at people with more chaotic lifestyles (e.g. those with mental health problems or those at risk of homelessness). Administrative databases may allow outcomes to be monitored over a long period without having to make direct contact with participants, but this is dependent on the type and quality of data collected.

Sample attrition is a problem partly because it reduces sample size (see above), but, more importantly, because it introduces a potential bias in the results. Those who drop out of the sample are unlikely to be a random sample of the target population. The longer the study, the greater this problem is likely to be. In the foster care study, for example, frequent movers, who are harder to trace, may contain a relatively high proportion of older and more troubled youngsters, with a history of placement breakdowns and poor relationships with their birth and foster families. The remaining sample are, therefore, likely to demonstrate better outcomes than perhaps would have been the case for the sample as a whole.

There is no easy way of correcting for the problem of sample attrition. At the very least,

the results should report the dropout rate and compare the baseline characteristics of those who dropped out of the sample with those who did not. If there are no significant differences, then this is probably sufficient. If not, then some attempt should be made to explore the possible implications; at the very least, this should identify the likely direction of any bias. In some studies with large samples and plenty of data, it may be possible to control for potential bias, by modelling the dropout process. An example of this was given in Box 12 (see Example 2).

### Follow-up

As well as monitoring outcomes at the beginning and end of an intervention, it is usually important to do at least one follow-up assessment. Some outcomes may emerge only in the long term; early intervention programmes for children, for example, have been shown to

have significant impacts on social and educational outcomes many years after the intervention is completed. Other outcomes are expected to have a more immediate impact, but it is still important to check whether any positive outcomes are sustained over time.

The timing of follow-ups is always slightly arbitrary and is often determined by practical considerations, such as the cost of tracking people over long periods of time. But, results can be very sensitive to timing (see, for example, the discussion of the postnatal support evaluation in the section on RCTs). Timing should, where possible, be informed by previous research. For example, there is evidence that people who have not reoffended for two years or more are far less likely to reoffend again, so interventions aimed at reducing reoffending should, ideally, monitor outcomes over a period of at least two years.

# 9 Analysing and interpreting results

## Analysis of results

### Combining costs and outcomes

When comparing two or more interventions with the same outcomes, data on costs and outcomes should be combined into some measure of cost-effectiveness.

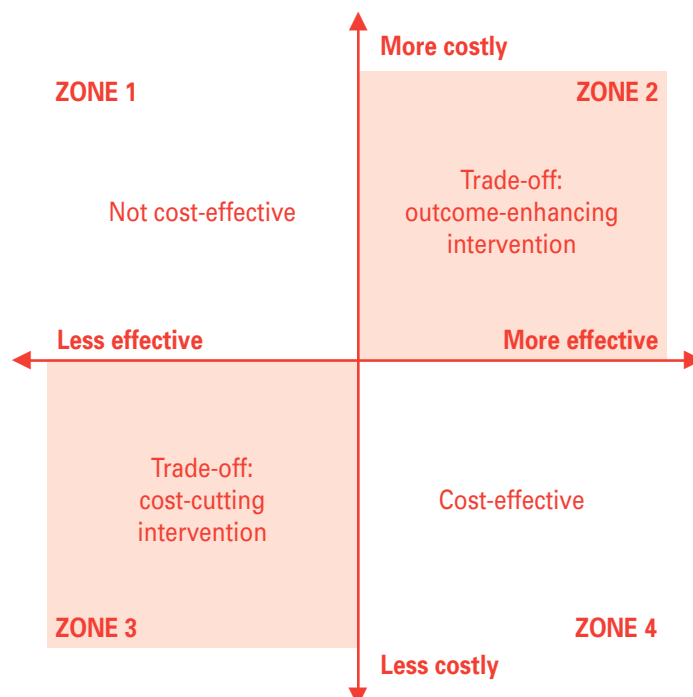
It is not possible to determine the cost-effectiveness of an intervention if there is more than one outcome, unless one intervention performs better, or equally well, on all counts and is no more expensive. In this situation, assessing relative cost-effectiveness is a matter of judgement for decision-makers, who need to decide how much weight to give to different outcomes. Economists can assist this judgement by presenting the results clearly and objectively.

Nor is cost-effectiveness analysis possible where outcome measures differ between the interventions being compared. If the outcomes

of two (or more) interventions are the same, but are measured in different ways, then it may be possible to compare their effect size,<sup>1</sup> although this should be done with care. This approach was used to make a crude comparison between the cost-effectiveness of two early reading interventions, which were evaluated in separate studies using different reading tests. The researchers concluded that a group-based intervention was more cost-effective than one-to-one tuition, because the mean effect size was comparable and costs were roughly ten times lower (Nicolson *et al.*, 1999).

Formal cost-effectiveness analysis is not essential if an intervention is found to dominate or be dominated by other options; this is the case if it is more effective and less costly than other interventions (i.e. Zone 4 in Figure 9), or less effective and more costly (i.e. Zone 1).

Figure 9 Assessing cost-effectiveness



However, where one intervention is found to produce greater benefit for greater cost or fewer benefits at lower cost (i.e. Zones 2 and 3 in Figure 9), then it is important to look at incremental cost-effectiveness (Drummond *et al.*, 1997).

Incremental analysis involves calculating the ratio of additional benefits to additional costs (i.e. how much more it costs to achieve a given improvement in outcomes, relative to the comparator intervention). This cost-effectiveness ratio can be compared to other interventions, employing the same outcome measure with preference given to those with the lowest cost per unit of outcome. Volume 2 provides more detailed guidance (Byford *et al.*, 2002, forthcoming).

### Cost and outcome variations

Costs and outcomes can vary substantially between individuals, especially for more flexible services, for example services that have adopted case management or systems that build in user choice. For example, in the first phase of the Maudsley Daily Living Programme, a community-based alternative to emergency admission to psychiatric hospital, costs varied by a factor of 27 between the least expensive and most expensive patients 12 to 20 months after entry to the study (Knapp *et al.*, 1994). It is, therefore, misleading to talk about the cost of treatment as if this were a uniquely defined figure.

Where data permits, it is important to explore these variations. Sub-group analysis looks at costs and outcomes for broad categories of client or user to see whether the intervention is more effective or cost-effective for certain groups of users. A scheme that is not cost-

effective for the general population may be found to be cost-effective for particular sub-groups. One criticism of the evaluation of a postnatal support scheme (see Example 2 in Box 8) was that it was not specifically designed to assess the scheme's effectiveness for more dependent mothers, who might be expected to benefit most from the extra support.

Cost or outcome functions are a more sophisticated technique for exploring variations in costs and outcomes between individuals, facilities, or areas. These use multiple regression techniques to look for statistical correlations between costs, outcomes, and user and provider characteristics. They are a good way of uncovering those factors that have a significant influence on costs or outcomes, especially when there are a lot of confounding factors. This is especially important when a study is not based on a randomised controlled trial, because differences between the intervention and comparison groups cannot be assumed to cancel each other out. By controlling for differences in the characteristics of clients, it is possible to isolate the net impact of the intervention being evaluated.

Cost and outcome functions are also useful for other purposes (Knapp, 1998), for example to examine whether a scheme is well targeted (i.e. whether expenditure on individuals is closely correlated with characteristics that indicate greater need) or to predict costs or outcomes from baseline information on user characteristics, which is useful for service planners.

### Other considerations

Economists are sometimes accused of focusing exclusively on efficiency at the expense of other criteria, such as equity (e.g. House, 2000). Studies



typically report overall or average outcomes without considering who stands to gain (or lose) most from the intervention. Equity considerations are likely to be important in the social welfare field and, at the very least, evaluation studies should examine the distributional effects on relevant sub-groups (e.g. by income group, age, severity of disability). It would be incorrect to say that economists have ignored the distributional effects of public policies, since many studies focus specifically on this issue (e.g. Blow and Crawford, 1997; Evans *et al.*, 1994; Hills, 1989). But, equity issues are not an integral part of most economic evaluation studies, despite strong recommendations in principle (e.g. Gold *et al.*, 1996).

In addition, a more sophisticated analysis might consider access to, or use of, services in relation to need or whether the way in which a service is delivered is equitable and perceived to be such. This analysis goes further than presenting the disaggregated results of the study and may involve more qualitative analysis – for example, in helping to understand how different people gain access to a particular service or how decisions are made about what services to offer to different users.

Some economists have attempted to build equity considerations into economic evaluation, by asking people to express their preferences for different equity outcomes (e.g. comparing a situation in which poorer households do slightly better, but outcomes across the whole population are slightly worse). However, a major study concluded that these techniques present significant, if not insurmountable, problems and recommended that information on distributional effects should instead be presented separately (Sassi *et al.*, 2001).

Other considerations, such as empowerment, fairness or justice, are concerned with the way services are delivered. These process issues can be important in their own right, quite independently of how they affect the achievement of programme objectives. The same objectives can be pursued in ways that either empower or disempower those affected, depending on how prescriptive the programme is and whether it is ‘user focused’.

Empowerment and other values should be included within the evaluation framework, where relevant, either as separate objectives or as constraints on the choice of available options. There is no reason why these values cannot be incorporated within the standard economic framework and there are already examples of this (e.g. Kendall and Knapp, 2000).

### Understanding how and why programmes work

Most economic evaluation is focused on the assessment of costs and outcomes. Typically, much less emphasis is given to understanding how interventions work and why they generate certain costs and outcomes. Some economists would see these as separate, and perhaps subsidiary, questions, which are not required to assess effectiveness or cost-effectiveness. Better studies will sometimes explore variations in costs and outcomes between different sub-groups and may even try to explain these variations, for example by looking for statistical associations between outcomes and certain characteristics of the programme. But, this does not usually get much beyond some tentative hypotheses, partly because sample size is rarely large enough to permit this type of statistical analysis.

There are good reasons why it is important to have a better understanding of how and why programmes work, particularly for more complex social welfare interventions. First, this can help to strengthen the validity of a study; if evaluators are able to show the links between a particular intervention and desired outcomes, then they can be more confident in attributing these outcomes to the intervention. This matters less if the study design is robust, since a well-conducted randomised controlled trial should ensure unbiased estimates of costs and outcomes. But, RCTs are not always feasible in the social welfare field for reasons discussed in the previous chapter. For less rigorous study designs, it cannot automatically be assumed that differences between the intervention group and the comparison group are due to the intervention being evaluated. In these situations, an evaluation will be more convincing if it can provide a plausible explanation as to how the programme led to positive outcomes for the intervention group. Box 18 describes the ‘theory of change’, which is one approach that makes questions about how an intervention works a central part of evaluation (see Example 1).

Second, an understanding of how and why interventions work can help to generalise the results of an evaluation, especially if outcomes are likely to be context-dependent. One criticism of experimental evaluation is that studies of the same intervention often produce very different results (Pawson and Tilley, 1997). This is not altogether unexpected given that people in various circumstances may respond differently to the same intervention or the intervention itself may vary in subtle ways, given different levels of training, competence and enthusiasm in different areas. (This is one reason why apparently

successful programmes often fail when they are ‘borrowed’ from elsewhere.) But, without an understanding of why the impact of a programme varies between places or population groups, the decision-maker is left none the wiser. If, on the other hand, the evaluation includes an analysis of how and why the programme is working, then it is easier to assess whether the same intervention is likely to be effective or cost-effective in other settings. It may be possible, for example, to identify certain features of the institutional setting that interact positively with the intervention and that need to be present to ensure successful outcomes. Alternatively, there may be certain characteristics that make some users more responsive to an intervention than others. ‘Realistic evaluation’ takes this a step further by arguing that the focus of evaluation should be on identifying effective policy mechanisms, rather than on measuring net outcomes (see Example 2 in Box 18).

More sub-group analyses would be useful and studies should be designed with this in mind, which means larger samples. But, economists should also be open to using other approaches to address these types of question as part of an economic evaluation. Although these approaches are usually presented as alternatives to experimental approaches, there is no reason in principle why they could not be combined, using experimental studies to obtain accurate estimates of outcomes and other approaches, such as Realistic Evaluation, to assess how far these results are likely to apply in different settings. There may also be some situations where a theory-based approach to evaluation could substitute for a more systematic assessment of costs and outcomes (see the example in Box 19).

**Box 18 Other approaches to evaluation**

*Example 1: theory of change*

The theory of change was developed as a tool for evaluating complex community-wide initiatives, where it has been argued that experimental approaches are not feasible (Connell, 1995). But, it has potential applications in other areas of social welfare. The theory of change (TOC) works from the assumption that programmes are based on explicit or implicit theories about how intended objectives are to be achieved. The initial stage of an evaluation involves working with key stakeholders to help define what the programme is aiming to achieve and by when and how they are going to get there. Thus, the evaluator starts by defining longer-term objectives and works backwards through the steps required to achieve them. Early stage or intermediate objectives are then established for each of these steps, so that the programme can be evaluated, and if necessary modified, at any stage.

Whereas longer-term milestones are more likely to focus on outcome measures, shorter-term milestones are more likely to be output or even process outcomes. Economists are often critical of process measures because they do not directly measure the achievement of end objectives. But, within a theoretical framework, such as TOC, process measures can be an effective way of assessing whether a programme's objectives are likely to be achieved (Hughes and Traynor, 2000). Further down the line, they provide a chain of evidence linking the intervention to the achievement of longer-term outcomes.

As an example, the ongoing evaluation of Healthy Living Centres is examining some of the theories of change that underlie different local initiatives. Some of these are seeking to reduce financial stress and the adverse impact this has on people's health. In the short term, the aim might be to set up a Credit Union in the area and encourage as many people as possible to join. In the medium term, this should encourage more people to save and would help to build up the community's capacity to address its own needs. In the longer term, this should help people to break out of debt and the anxiety that goes with this, as well as increasing people's trust in, and solidarity with, their local community. This, it is hoped, would eventually lead to better health outcomes for the most disadvantaged, which is the primary objective of Healthy Living Centres.

*Example 2: realistic evaluation*

Proponents of realistic evaluation argue that evaluation should not focus on measuring the net outcomes of programmes, which are unlikely to tell you very much about the effectiveness of the same programme in other settings (Pawson and Tilley, 1997). Instead, it should seek to identify effective mechanisms that lead to desirable outcomes. For example, the use of CCTVs in car parks may involve one or more of the following mechanisms:

*continued overleaf*

- ‘You’ve been framed’ mechanism: deterring potential offenders, by increasing the risk of being caught.
- ‘Effective deployment’ mechanism: facilitating the deployment of security staff and police to areas where suspicious behaviour is occurring.
- ‘Memory jogging’ mechanism: reminding drivers to take greater care (e.g. to remember to lock their cars and operate any security devices).
- ‘Appeal to cautious’ mechanism: providing a haven for safety-conscious drivers.

Because the success of many interventions is context dependent, a realistic evaluation should also specify the context in which these different mechanisms are likely to be triggered. In the above example, the ‘effective deployment’ mechanism is more credible in city car parks with greater security presence than in areas where car parks are more widely dispersed. The success of the ‘memory jogging’ mechanism, on the other hand, will depend on the ‘state-of-the-art’ of security devices (e.g. their prevalence, effectiveness and ease of use).

These context–mechanism–outcome configurations are developed by evaluators with the help of key stakeholders, including those responsible for designing and implementing the programme. They start off as theories or hypotheses, but are then tested against all the available evidence. For example, police and security firms could be interviewed to see if the installation of CCTVs has altered the way they deal with car-related crime and police statistics might be used to assess whether there has been an improvement in detection rates. Alternatively, a survey of drivers might be used to examine any impact on their behaviour (e.g. whether the presence of CCTV cameras had an effect on their choice of car park).

Realistic evaluation is not prescriptive about research methods, but encourages a mix of qualitative and quantitative approaches. Its purpose is to identify effective mechanisms, the context in which these are likely to be triggered and expected outcomes under these circumstances. This information, it is argued, is more transferable, and therefore more useful to decision-makers, than one-off results from experimental evaluations. If, in our example, the ‘effective deployment’ mechanism were found to be the dominant mechanism, then policy-makers would know that the use of CCTVs in car parks is unlikely to be effective in more rural areas without an adequate police or security presence; they would also know that, even in urban areas, the use of CCTV is likely to be effective only if it is coordinated with the deployment policies of local police and security firms.

### Box 19 Example of theory-based economic evaluation

#### *Evaluation of GP fundholding*

The aim of the fundholding initiative was to improve the efficiency of hospital services by giving GP practices direct responsibility for purchasing certain hospital services, including diagnostic tests and low-cost surgical treatments, thus creating an internal market within the NHS. This initiative could have been evaluated by comparing costs and health outcomes for patients of fundholders and non-fundholders, but this would have been very expensive and would not have provided any answers in the short term. Instead, this study was based around a set of specific evaluation questions, which were informed by initial hopes or concerns about the scheme and by micro-economic theory. For example:

- Whether fundholding would improve services for patients: micro-economic theory predicts that giving purchasing power to GPs should encourage competition between nearby hospitals, leading to better and/or cheaper services.
- Whether it would alter prescribing patterns: under the old system, GPs had no financial incentive to prescribe 'rationally', because these costs were paid by a third party. Policy-makers were hoping that, because fundholders would now have an incentive to control their prescribing behaviour, this would lead to a reduction in prescriptions and a switch to more cost-effective medicines.
- Whether fundholders would manage to keep within their budgets: one of the Government's concerns was that GPs might run out of money before the year end and that they would have to step in to ensure patients did not go without treatment.
- Whether fundholders would seek to manipulate their patient list: the Government was also worried that fundholders might try to 'cream-skim' the cheapest patients in order to earn a larger surplus for their practice, using information from patients' medical records to exclude those who were likely to be the heaviest users of health services.

Each of these questions was addressed using a range of different research methods, including regular interviews with GPs (to understand their motivations for becoming fundholders and the impact this was having on their practice), financial and prescribing data (to monitor their expenditure against budget and to compare the prescribing patterns of fundholders and non-fundholders) and an analysis of a sub-sample of patients' medical records (to see whether this information could be used to identify potentially expensive patients).

The authors concluded that GP fundholding was effective in promoting efficiency to a significant, though not 'revolutionary', extent and that the Government's concerns about budget management were unfounded. While they showed the incentives were there for 'cream-skimming', they found no evidence that this was happening in practice. On this basis, they argued that the benefits of fundholding were likely to outweigh the additional administrative costs.

(Source: Glennerster *et al.*, 1994)

### Interpretation and presentation of results

The results of economic evaluations are often presented poorly or in a misleading way. While quantitative results can be very powerful, they are also easy to misinterpret. In the studies we looked at in our review (see Chapter 3), there seemed to be a tendency to exaggerate the impact of the interventions concerned. In some cases, the discussion at the end of the paper was very uncritical, focusing only on the positive aspects of the scheme and drawing conclusions that went beyond what the evidence seemed to support. Results that were ambiguous were sometimes presented as supporting the expansion of a particular programme. In other cases, insufficient information was presented to be able to make a judgement about the reliability of the results. One of the main deficiencies was the lack of attention given to exploring alternative explanations for the results.

In the health care field, there are fairly clear guidelines for presenting the results of economic evaluations, which seek to ensure that certain standards are met in all studies and that information on key aspects of the evaluation design are clearly laid out (e.g. Eisenberg, 1989; Siegel *et al.*, 1997; Task Force on Principles for Economic Analysis of Health Care Technology, 1995). This is beginning to lead to greater uniformity in the way results are presented (though it may also be stifling innovation).

It is probably not appropriate to apply the same set of criteria across all areas of social welfare. The guidelines developed in the health care field are too rigid for most other areas of social welfare, so it may be better to think in terms of a set of common principles that studies

should attempt to adhere to. Good practice would include the following:

- Setting results in the context of previous studies of similar interventions: comparing the results, highlighting any differences and seeking to explain these.
- Acknowledging any weaknesses in the design of the study and identifying any potential sources of bias that may result from these. Evaluators should make some attempt to assess the direction and significance of any bias (e.g. by examining, and if possible controlling for, differences in the characteristics of the intervention and comparison groups). This may have to be done more qualitatively, especially if sample size is small.
- Exploring, and if possible ruling out, alternative explanations for the results.
- Testing whether the results are consistent with a priori expectations and attempting to explain any unusual or unexpected results.
- Considering the significance of outcomes and the range of uncertainty (e.g. by estimating confidence intervals and presenting the results of sensitivity analysis on key parameters).
- Being transparent about the quality and quantity of the data and how this might affect the robustness of the results (e.g. the validity of outcome measures).
- Being clear about what is (and is not) included in the final analysis (e.g. which

benefits have been monetised and included in a cost-benefit analysis) and considering the implications of omitting some of these effects.

- Clearly stating and justifying any assumptions underlying the analysis.

A more comprehensive checklist for assessing economic evaluations is included in Volume II of this report (Byford *et al.*, 2002, forthcoming).

Presentation of qualitative results is equally important. It is probably easier than with quantitative data to present a distorted picture by using material selectively to support a

particular conclusion, although systematic approaches to analysing qualitative data are available (Pope *et al.*, 2000). It is also harder to present the results of qualitative analysis concisely, which is both an advantage and a disadvantage. On the one hand, it means that qualitative results will tend to have less of an immediate impact on the reader. On the other hand, there is an opportunity to present a more complex and subtle message. In the case of economic evaluation, it is likely that most qualitative analysis will be used to support or qualify the results of a quantitative study, so the two should, ideally, be integrated in the final report.

# 10 Managing the evaluation process

The circumstances in which evaluation takes place can be very important in shaping the approach to evaluation. Our case studies were particularly helpful in highlighting the importance of these issues.

## Involving key stakeholders

One of the lessons from other evaluators is that it is useful, and sometimes necessary, to involve stakeholders more closely in the evaluation process. This includes those responsible for funding, designing or administering a programme, as well as service users.

For some types of economic evaluation, including cost-benefit and cost-utility analysis, stakeholder involvement is crucial in eliciting people's preferences for different outcomes. Arguably, this form of outcome valuation is the 'ultimate' stakeholder approach, since it is based on the preferences of those most directly affected by the intervention. The techniques economists use to value outcomes were described in the section on measuring outcomes (see Chapter 7). Some economists argue that the views of users (and the wider public) are the only ones that ought to count, because they receive the benefits and ultimately bear the cost (as taxpayers).

However, economists rarely involve stakeholders in other aspects of evaluation and are reluctant to involve those with a vested interest in the programme (other than users) in case this should threaten the objectivity of the study. By contrast, evaluators with a social policy background will typically involve a wide range of stakeholders at various stages in the evaluation process:

- helping to clarify the purpose of the evaluation study
- clarifying the objectives that are driving the operation of the programme
- identifying possible unintended or knock-on effects of a programme
- helping to define those outcome or cost dimensions that are of most importance
- gaining a better understanding of how a programme is supposed to operate and how it is operating in practice.

Thus, stakeholders can be involved in ways that many economists would not normally consider. Their input is likely to be particularly valuable in certain situations. For example, if the programme's objectives are not clearly defined or the programme is a complex one, then it is even more important to involve stakeholders in the process of clarifying objectives.

In some areas of social welfare, such as community development, stakeholders are actively encouraged to be part of the evaluation process. One of the fundamental aims of community development is to 'empower' local people by developing their skills and knowledge. Evaluation, if properly conducted, can be part of this learning process, giving people an opportunity to reflect and build on their experiences (Barr and Hashagen, 2000). Giving stakeholders more 'ownership' of the evaluation may also facilitate the evaluation process (e.g. by improving access to information) and increase the likelihood that the study's recommendations will be acted upon (Patton, 1996).



But, there are also potential problems with involving stakeholders. Some stakeholders may be reluctant to cooperate with an evaluation if they feel it is imposed on them or they are too busy. Economic evaluation, in particular, is sometimes perceived to be part of a cost-cutting exercise and people may feel threatened by this, so it is important to explain the purpose of the study and why their involvement is important. Another danger is that evaluators may align themselves too closely with a particular stakeholder whose views may not be representative.

### Working with other disciplines

Economists often work in parallel with other evaluators, rather than as part of a truly multi-disciplinary team. They may be assigned a particular role, usually a costing study, and carry this out more or less independently of the rest of the evaluation team. However, there are good reasons why it is desirable, if not essential, to build bridges between economists and other disciplines:

- to help dispel misconceptions (e.g. about what economic evaluation is really about)
- to establish a common understanding of concepts, overcoming some of the barriers created by the use of technical jargon
- to create opportunities for learning from other disciplines: for example, economists have a lot to learn about the importance of context and process
- to encourage a more integrated approach to evaluation (e.g. evaluating costs and outcomes together)

- to give economists a more realistic view of what can be achieved through a better understanding of the complexities of the programme
- to give other evaluators a better understanding of what economists can contribute at different stages in the evaluation process.

However, experience from the case studies and other evaluations being undertaken by members of the research team suggests a number of barriers to effective collaboration between economists and other evaluators. A deeper understanding of these barriers, on the part of different disciplines, can help researchers to overcome these problems at the outset of an evaluation:

- *Misunderstanding of the purpose of economic evaluation:* the involvement of economists is often seen to be part of a cost-cutting exercise, whereas economic evaluation is about the efficient use of resources and so could equally be used to justify higher spending on cost-effective services.
- *Conflict over issues of design:* the contrast between many economists' preference for large randomised control trials (RCTs) and many social researchers' preference for in-depth, qualitative research can generate significant problems for multi-disciplinary research. On a number of occasions, members of our team have been advised to withdraw from studies where a decision has been made not to randomise between groups. Had such advice been taken, progress in this area would not have been made. Economists

need to be more sympathetic to the substantial difficulties of carrying out RCTs in the social welfare field. Similarly, researchers more familiar with qualitative methods need to have a greater appreciation of the benefits of quantitative research. The relative merits of different approaches should be discussed in an attempt to find the most suitable compromise. Such discussions have led to the design of a number of studies comprising large quantitative samples from which smaller samples have been selected for qualitative exploration.

- *Conflict over the use of established outcome scales:* some disciplines are sceptical of the use of standardised outcome scales, often because the scales do not seem to measure exactly what the researcher wishes to explore. Instead, the preference is to design new measures that more accurately relate to the objectives of interventions. This is a reasonable criticism of outcome scales, but the advantages of such measures should not be dismissed lightly. In order to provide robust estimates of cost-effectiveness and allow comparison across a broader spectrum of interventions, established outcome measures are valuable tools. Once again, researchers should explore the pros and cons of both standardised scales and 'home-made' measures, and ideally attempt to incorporate both.
- *Lack of understanding of qualitative methods on the part of economists:* qualitative research is an important research tool that

plays a fundamental role in the development of evidence and policy in many areas of social welfare. Economists, however, are often unaware of the methods and uses of qualitative research and may not recognise how qualitative analysis can support their own research. Economists should be open to using other research methods with the help of qualitative researchers.

- *Conflict over the resources devoted to collecting economic data:* economic evaluation is commonly seen as an 'add-on' component to the 'more important' task of evaluating outcomes. It is not unusual for economists to find that they are expected to do their work within the conditions dictated to them by the outcomes research and for a relatively small share of the resources available for evaluation, including funding, time and even questionnaire space. It is important for other evaluators to understand that significant constraints on the economic component can damage the usefulness of the whole study.
- *Desire to keep the economic component of a study separate:* there is a tendency for the economic evaluation to be viewed as distinct from the outcome component of a study, such that the economist is often expected to work alone and solely on the assessment of costs. This undermines the economic evaluation which requires both cost and outcome data in a suitable form; it also limits the potential benefits from multi-disciplinary research.

To improve the current situation, it is necessary to build genuinely multi-disciplinary teams, bringing together economists, social researchers and other specialists (e.g. psychiatrists if evaluating a mental health initiative). Other evaluators need to involve economists at an earlier stage in the evaluation process, while economists should be more willing to 'get their hands dirty', for example by visiting projects or attending research interviews. In addition, both sides need to be willing to use a wider range of research methods. Economists would benefit from training in the use of qualitative research methods or simply from observing the way these methods are used in practice. But, what is needed, above all else, is a genuine understanding and appreciation of what different specialists have to offer to evaluation.

### Constraints on evaluation

Evaluation is always carried out under various constraints, which can have a significant impact on the design of an evaluation study. The most obvious constraint is the availability of resources for evaluation. Although most public and voluntary bodies now acknowledge the need to evaluate their own programmes, resources are not always made available to do this properly. In the case of the Safe In The City programme, in the past no resources had been specifically earmarked for evaluation (as opposed to internal monitoring) and reliance had been put on external funding for the main evaluation study.

Lack of resources for evaluation may limit sample size, the amount of data that can be collected and the time frame. From a practical

point of view, it is difficult to obtain funding for long-term studies, even though some programmes are expected to have an impact only in the longer term. In some cases, the cost of carrying out a proper experimental study may be difficult to justify given the size of the programme, in which case the aims of the evaluation need to be scaled back accordingly. But, in other cases, programme funders and policy-makers should consider making more resources available for evaluation. It is not sufficient to acknowledge the importance of rigorous evaluation if the necessary funding is not built into programme budgets. This in turn requires a better understanding of the cost of good quality economic evaluation, which varies widely. Table 3 outlines the main factors affecting the cost of these studies. When resources are constrained, there is a trade-off between the comprehensiveness of data collected and the cost and time required to collect such data. Interviewer-based instruments are likely to provide the best data but are relatively expensive. Administrative data are cheap, but may not be ideal for evaluation purposes.

Other constraints on evaluators are imposed by the design of the programme, over which the evaluator may have little control, including:

- Decisions over who participates in the programme: classic experimental studies require that some potential clients are randomly assigned to a control group, but those administering the programme may not want to deny people access to the scheme or may wish to use other criteria to target their programme.

**Table 3 Factors affecting the cost of economic evaluation**

Factor	Explanation
Study design	Randomised control trials are often (but not always) more expensive than other designs, because of the emphasis on many of the factors listed below (e.g. large sample sizes, prospective data collection and face-to-face interviews).
Sample size	Dependent on the source of data (see below), costs will usually increase with sample size, because more interviews need to be carried out, more data needs to be entered and a longer recruitment period is needed.
Source of data	Administrative databases are the cheapest source, while face-to-face interviews are the most expensive. Other sources, such as postal questionnaires and records kept by service providers, lie somewhere in between, although the latter can be time-consuming to analyse and, therefore, expensive.
Length and frequency of follow-ups	Greater frequency increases data collection costs. Length of the follow-up period also increases costs, because researchers need to be employed for longer and respondents may be harder to trace.
Retrospective or prospective data collection	Retrospective evaluation, though less desirable, is generally cheaper than prospective evaluation, because it relies on existing data.
Need for special interviews or data collection tools	Designing new questionnaires, though often necessary, can be time-consuming, especially if these need to be piloted.
Type of economic evaluation	It is difficult to make generalisations, but cost-benefit analysis will tend to be more expensive if a special survey is needed to put monetary values on outcomes.
Ethical approval	Gaining ethical approval for an evaluation is a time-consuming and administratively expensive process. This is not normally required outside health care, but may become more of an issue in future.
Whether have to estimate own unit costs	In some cases, it is possible and appropriate to use existing unit cost figures (e.g. national averages), which avoids the time and costs involved in collecting micro-level resource data for specific services.
Need for data entry	Data entry adds significantly to the cost of an evaluation study, unless existing data sources can be exploited (or data are entered by service providers or even users).
Number of data collection points	Cost will be greater if data collection is more dispersed (e.g. from many different agencies, different parts of the country, or different countries).
Analysis of data	Data from randomised controlled trials are relatively easy to analyse. Quasi-experimental studies, on the other hand, may involve complex statistical techniques, requiring specialist input, which can be expensive.

---

- Changes in policy or the natural development of programmes over time may adversely impact on the evaluation, substantially altering the nature of the intervention that is being evaluated (or the comparator intervention). A common problem is that the services offered to the comparison group may start to imitate those provided to the intervention group, especially if the intervention is ‘in fashion’.
- If the programme is very small, this may restrict the achievable sample size and reduce the statistical power of a quantitative study. It is also likely to restrict the amount of funding available for evaluation.

In addition, evaluators operate under certain constraints imposed by those funding the study. They have a vested interest in the results and may use their position to influence the way the study is carried out. Sponsors may also impose artificial time constraints on the evaluation, perhaps because they want early results for political reasons. Academic researchers face their own pressures to produce significant or unexpected results, since these are more likely to be published in academic journals (e.g.

Easterbrook *et al.*, 1991).

Other constraints are specific to economists, because of their role in the evaluation team. Economists may only be involved at a late stage and, even when they become involved, their role may be restricted to certain aspects of evaluation. An example of this is a recent study of the financial cost of childhood conduct problems (Scott *et al.*, 2001). This economic analysis was not conducted alongside the social/clinical evaluation and there was no input by economists into the design of the original questionnaire on which their cost estimates had to be based. In conducting a retrospective analysis, many costs and resource utilisation data were not available (e.g. on the use of social services, voluntary sector costs and caregiver costs). As a result, the researchers were able to produce only a partial (and conservative) estimate of the long-term costs.

Finally, economic evaluation is harder when similar interventions have not been evaluated in this way before. Suitable outcome scales may not be available, unit cost figures cannot be ‘borrowed’, the required sample size is more difficult to judge and the results are not always easy to interpret without other studies to compare them with.

# 11 Summary and conclusions

## Summary

Greater and better use of economic evaluation is needed in the social welfare field to help inform decisions about the use of limited resources, alongside other political, ethical and pragmatic considerations. To date, little economic evaluation of social welfare interventions has been carried out in the UK, outside health care. Most of the studies we identified are from the US and predominantly in areas closely related to health care. The quality of these studies is very variable.

There are good reasons why economic evaluation has been slow to develop. Demand for economic evaluation has been lacking and indiscriminating, and there has also been resistance from other evaluators or practitioners. In addition, many social welfare interventions are very complex, being characterised by a high degree of user involvement, significant variability within the programme, complex and long-term outcomes and multi-agency involvement. These complexities undoubtedly pose significant challenges, but they should not normally preclude the achievement of good quality economic evaluation.

Economic evaluation should be part of a broader evaluation strategy that seeks to define, refine and understand a programme before attempting a more systematic evaluation of its effectiveness and cost-effectiveness. Economists have an important contribution to make at all stages of the evaluation, because of the particular perspective they bring, and early involvement would help to prepare the ground for a full economic evaluation at a later stage, for example by ensuring the right data are collected. Funders and decision-makers need to be clear about what stage they think they are at

in a programme's development, since this will determine the priorities for evaluation at any given point and what can realistically be expected of the evaluation.

The first priority for any evaluation is to clarify a programme's objectives. This is likely to be a fundamental and substantive element of the evaluation of social welfare interventions, given that their objectives are often multiple, ill defined and may vary over time and between stakeholders. The process of identifying and assessing relevant costs and outcomes is also a critical aspect of evaluation. The cost side, in particular, needs to be given more emphasis than at present and to be treated as an integral part of the evaluation process. Currently, most economic evaluations adopt a narrow perspective on costs, frequently ignoring non-monetary costs and knock-on effects on the use of other services.

It is both desirable and feasible to conduct more randomised controlled experiments of social welfare interventions. Some of the concerns about this approach are either mistaken or exaggerated and can be overcome with careful design and the use of complementary research methods. Where random assignment is not feasible either for practical or ethical reasons, the priority should be to make the evaluation design as robust as possible. Creative ways can be found to achieve 'near-randomisation'; statistical techniques can be used to control for differences between the intervention and comparison groups; validated scales can be used to measure outcomes; and multiple outcomes and extended follow-up can help to provide a more comprehensive assessment of a programme's impact. In some cases, the programme itself can be designed to

make it easier to evaluate, for example by piloting or staggering the implementation of the programme over time, in order to create ‘in-built’ comparators.

The design of evaluations in the social welfare field will frequently fall short of the ‘ideal’. In general, the more complex an intervention, the more likely it is that an experimental approach to economic evaluation will be inadequate on its own. There is a greater role for qualitative research methods, in particular, in helping to address many of the limitations of experimental approaches, for example to check the appropriateness of outcome measures and to corroborate the findings from the quantitative analysis.

The results of economic evaluations need to be presented and interpreted more carefully than is often the case, acknowledging weaknesses in the study design, exploring potential biases that may result, and being clear about the range of uncertainty surrounding estimates of costs and outcomes. There are also good reasons why economists should pay more attention to how and why programmes work, especially for more complex programmes. This can help strengthen a study’s validity, by specifying the links between the intervention and observed outcomes, and can improve its generalisability, through a better understanding of the context in which a particular intervention is likely to be effective.

These suggestions for adapting or building on current approaches to economic evaluation may require economists to take on a different role to the one they are accustomed to. Economists jealously guard their objectivity, yet the evaluation of social welfare programmes can rarely be carried out properly without involving

stakeholders in the evaluation process, including those responsible for designing the programme, practitioners who deliver the programme and service users. This may be necessary early on in helping to clarify the objectives of the programme and in identifying possible unintended effects of the programme and, at a later stage, in assessing outcomes and understanding how these outcomes were generated. Traditionally, other evaluators have undertaken many of these evaluation activities, but we do not think it is advisable to separate purely economic components from other aspects of evaluation. This means learning to work more closely with evaluators from other disciplines, respecting and learning from each other’s perspectives.

Economic evaluation is nearly always carried out under various constraints that will substantially affect what can be done in practice, including limited resources, time constraints, ‘political’ pressures, lack of control over the programme and lack of suitable research instruments. Limited funding is a major obstacle to more rigorous approaches to economic evaluation and something that needs to be addressed by those responsible for commissioning evaluation studies.

## Conclusions

Economists have an important contribution to make to evidence-based policy-making and the demands for more rigorous evidence on the effectiveness and cost-effectiveness of different interventions are increasing. If done well, economic evaluation can provide valuable information to decision-makers faced with difficult choices about how to target scarce

resources. Although good quality economic evaluation is still rare in many areas of social welfare, there are examples of good practice, which demonstrate that rigorous economic evaluation can be carried out across a wide range of social welfare interventions.

The complex nature of many social welfare interventions, however, makes economic evaluation more challenging. Conventional approaches to economic evaluation in health care provide a good starting point, but they need to be adapted to reflect these additional

complexities. This requires a pragmatic approach, recognising that compromises may have to be made in the design of evaluation studies and that other research methods need to be used appropriately to help improve the validity and generalisability of the results. There is considerable scope for economists to learn from other perspectives on evaluation. But, economists also have plenty to offer in return, provided they are willing to embrace a more flexible and eclectic approach to economic evaluation.



# Note

## Chapter 9

- 1 By dividing the net effect by the standard deviation, this adjusts for differences in the scaling of different outcome measures.

# References

- Bagley, C. and Pritchard, C. (1998) 'The reduction of problem behaviours in at-risk youth: an experimental study of school social work with cost-benefit analyses', *Child and Family Social Work*, Vol. 4, No. 4, pp. 219–26
- Barr, A. and Hashagen, P. (2000) *ABCD Handbook: A Framework for Evaluating Community Development*. London: Community Development Foundation
- Beck, E.J., Beecham, J., Mandalia, S., Griffith, R., Walters, M.D., Boulton, M. and Miller, D.L. (1999) 'What is the cost of getting the price wrong?', *Journal of Public Health Medicine*, Vol. 21, No. 3, pp. 311–17
- Beecham, J. (2000) *Unit Costs not Exactly Child's Play: A Guide to Estimating Unit Costs for Children's Social Care*. London: Department of Health
- Beecham, J. and Knapp, M. (1992) 'Costing psychiatric interventions', in G. Thornicroft, C. Brewin and J. Wing (eds) *Measuring Mental Health Needs*. London: Gaskell
- Blow, L. and Crawford, I. (1997) *The Distributional Effects of Taxes on Private Motoring*. London: Institute of Fiscal Studies
- Briggs, A. (2000) 'Economic evaluation and clinical trials: size matters', *British Medical Journal*, Vol. 321, pp. 1362–3
- Britten, N. (1995) 'Qualitative research: qualitative interviews in medical research', *British Medical Journal*, Vol. 311, pp. 251–3
- Buxton, M.J., Drummond, M., Van Hout, B., Prince, R., Sheldon, T., Szucs, T. and Vray, M. (1997) 'Modelling in economic evaluation: an unavoidable fact of life', *Health Economics*, Vol. 6, pp. 217–27
- Byford, S., Harrington, R., Togerson, D., Kerfoot, M., Dyer, E., Harrington, V., Woodham, A., Gill, J. and McNiven, F. (1999) 'Cost-effectiveness analysis of a home-based social work intervention for children and adolescents who have deliberately poisoned themselves: the results of a randomised controlled trial', *British Journal of Psychiatry*, Vol. 174, pp. 56–62
- Byford, S., McDaid, D. and Sefton, T. (2002, forthcoming) *Economic Evaluation in the Social Welfare Field: How to Make the Most of it*. York: Joseph Rowntree Foundation
- Centre for Management and Policy Studies (2001) *Modernising Policy Development: Nine Features of Modern Policy Making*. <http://www.cmps.gov.uk/> (accessed in October 2001)
- Challis, D., Darton, R., Johnson, L., Stone, M. and Traske, K. (1995) *Care Management and Health Care of Older People: The Darlington Community Care Project*. Aldershot: Ashgate
- Clinch, P.J. and Healy, J.D. (2001) 'Cost-benefit analysis of domestic energy efficiency', *Energy Policy*, Vol. 29, pp. 113–24
- Coast, J. (1999) 'The appropriate uses of qualitative methods in health economics', *Health Economics*, Vol. 8, pp. 345–53

- Connell, J.P. (ed.) (1995) *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts*. Queenstown, MD: Aspen Institute
- Davies, B. and Challis, D. (1986) *Matching Resources to Needs in Community Care: An Evaluated Demonstration of a Long-term Care Model*. Aldershot: Gower
- Department of the Environment, Transport and the Regions (2000) *Fuel Poverty: The New HEES*. London: The Stationery Office
- Department of Health (2000) *Framework for the Assessment of Children in Need and their Families: The Family Pack of Questionnaires and Scales*. London: The Stationery Office
- Deschenes, E. and Greenwood, P. (1998) 'Alternative placements for juvenile offenders: results from the evaluation of the Nokomis Challenge Program', *Journal of Research in Crime and Delinquency*, Vol. 35, No. 3, pp. 267–94
- Drummond, M.F., O'Brien, B., Stoddart, G.L. and Torrance, G.W. (1997) *Methods for the Economic Evaluation of Health Care Programmes*, 2nd edn. Oxford: Oxford Medical Publications
- Easterbrook, P., Berlin, J., Gopalan, R. and Matthews, D. (1991) 'Publication bias in clinical research', *The Lancet*, Vol. 337, No. 8746, pp. 867–72
- Eisenberg, J. (1989) 'Clinical economics: a guide to the economic analysis of clinical practices', *Journal of the American Medical Association*, Vol. 262, No. 20, pp. 2879–86
- Ekblom, P., Law, H. and Sutton, M. (1996) *Safer Cities and Domestic Burglary*. London: Home Office
- Elixhauser, A., Halpern, M., Schmier, J. and Luce, B. (1993) 'Health care cost-benefit and cost-effectiveness analysis (CBA/CEA) from 1979 to 1990: a bibliography', *Medical Care*, Vol. 31, No. 7, Supplement
- Elixhauser, A., Halpern, M., Schmier, J. and Luce, B. (1998) 'Health care CBA and CEA from 1991 to 1996: an updated bibliography', *Medical Care*, Vol. 36, No. 5, Supplement
- EuroQol Group (1990) 'EuroQol – a new facility for the measurement of health-related quality of life', *Health Policy*, Vol. 16, pp. 199–208
- Evans, M., Piachaud, D. and Sutherland, H. (1994) *Designed for the Poor – Poorer by Design?: The Effects of the 1986 Social Security Act on Family Incomes*. Welfare State Programme Discussion Paper No. 105. London: London School of Economics
- Farkas, G. (1998) 'Reading one-to-one: an intensive program for serving a great many students whilst still achieving large effects', in J. Crane (ed.) *Social Programs that Work*. New York: Russell Sage Foundation
- Forrester, D., Frenz, S., O'Connell, M. and Pease, K. (1990) *The Kirkholt Burglary Prevention Project: Phase II*. London: Home Office
- Gardiner, K. (1997) *Bridges from Benefit to Work: A Review*. York: YPS for the Joseph Rowntree Foundation
- Glennerster, H. and Turner, T. (1993) *Estate Based Housing Management: An Evaluation*. Report for the Department of the Environment. London: HMSO

- Glennester, H., Matasaganis, M., Owen, P. and Hancock, S. (1994) *Implementing GP Fundholding: Wild Card or Winning Hand?* London/Bristol, PA: Open University Press
- Gold, M., Siegel, J., Russell, L. and Weinstein, M. (1996) *Cost-effectiveness in Health and Medicine*. New York: Oxford University Press
- Greenwood, D.J. (1998) *Introduction to Action Research: Social Research for Social Change*. Thousand Oaks, CA: Sage
- Guba, E.G. and Lincoln, Y.S. (1989) *Fourth Generation Evaluation*. London: Sage
- Hasluck, C., McKnight, A. and Elias, P. (2000) *Evaluation of the New Deal for Lone Parents: Early Lessons from the Phase One Prototype – Cost-benefit and Econometric Analyses*. Department of Social Security Research Report No. 110. Huddersfield: The Charlesworth Group
- Hausman, D. and McPherson, M. (1994) 'Economics, rationality, and ethics', in D. Hausman (ed.) *The Philosophy of Economics: An Anthology*, 2nd edn. Cambridge: Cambridge University Press
- Healey, A. (2001) 'An economic approach to assessing the impact of violent criminal offending on individual well being: results from exploratory survey of the general public', presentation at conference organised by the Institute of Psychiatry, 'Economics of mental health', 12 October
- Hills, J. (1989) *Distributional Effects of Housing Subsidies in the United Kingdom*. Welfare State Programme Discussion Paper No. 44. London: STICERD
- HM Treasury (1997) *Appraisal and Evaluation in Central Government*. London: The Stationery Office
- Holtermann, S. (1995) *Investing in Young Children: A Reassessment of the Cost of an Education and Day Care Service*. London: National Children's Bureau
- Holtermann, S. (1998) *Weighing it up: Applying Economic Evaluations to Social Welfare Programmes*. York: YPS for the Joseph Rowntree Foundation
- House, E.R. (2000) 'The limits of cost-benefit analysis', *Evaluation: The International Journal of Theory, Research and Practice*, Vol. 6, No. 1, pp. 79–86
- Hughes, M. and Traynor, T. (2000) 'Reconciling process and outcome in evaluating community initiatives', *Evaluation: The International Journal of Theory, Research and Practice*, Vol. 6, No. 1, pp. 37–49
- Hutton, J. (1992) 'Cost-benefit analysis in health care expenditure decision-making', *Health Economics*, Vol. 1, pp. 213–16
- Jan, S. (1998) 'A holistic approach to the economic evaluation of health programs using institutionalist methodology', *Social Science and Medicine*, Vol. 47, No. 10, pp. 1565–72
- Johannesson, M. (1995) 'A note on the depreciation of the societal perspective in economic evaluation of health care', *Health Policy*, Vol. 33, pp. 59–66
- Johannesson, M., Joensson, B. and Karlsson, G. (1996) 'Outcome measurement in economic evaluation', *Health Economics*, Vol. 5, No. 4, pp. 279–96

- Joseph Rowntree Foundation (2001a) *Current Priorities*. <http://www.jrf.org.uk/funding/priorities/> (accessed in October 2001)
- Joseph Rowntree Foundation (2001b) *Specific Requirements of the Committee*. <http://www.jrf.org.uk/funding/priorities/scdc.asp> (accessed in October 2001)
- Keen, J. and Packwood, T. (1995) 'Qualitative research: case study evaluation', *British Medical Journal*, Vol. 311, pp. 411–46
- Kendall, J. and Knapp, M. (2000) 'Measuring the performance of voluntary organisations', *Public Management*, Vol. 2, No. 1, pp. 105–32
- Kitzinger, J. (1995) 'Qualitative research: introducing focus groups', *British Medical Journal*, Vol. 311, pp. 299–302
- Knapp, M. (1998) 'Making music out of noise – the cost function approach to evaluation', *British Journal of Psychiatry*, Vol. 173, Supplement 36, pp. 7–11
- Knapp, M. (1999) 'Economic evaluation and mental health: sparse past ... fertile future?', *The Journal of Mental Health Policy and Economics*, Vol. 2, pp. 163–7
- Knapp, M., Cambridge, P., Thomason, C., Beecham, J., Allen, C. and Darton, R. (1992) *Care in the Community: Challenge and Demonstration*. Aldershot: Ashgate.
- Knapp, M., Beecham, J. and Koutsogeorgopoulou, V. *et al.* (1994) 'Service use and costs of home-based versus hospital-based care for people with serious mental illness', *British Journal of Psychiatry*, Vol. 165, pp. 195–203
- Kushner, Saville (1996) 'The limits of constructivism', *Evaluation: The International Journal of Theory, Research and Practice*, Vol. 2, No. 2, pp. 189–200
- Long, C., Williams, M. and Hollin, C. (1998) 'Treating alcohol problems: a study of programme effectiveness and cost effectiveness according to length and delivery of treatment', *Addiction*, Vol. 93, No. 4, pp. 561–71
- McDaid, D. (2001) 'Estimating the costs of informal care for people with Alzheimer's Disease: methodological and practical challenges', *International Journal of Geriatric Psychiatry*, Vol. 16, pp. 400–5
- McDaid, D., Sefton, T. and Byford, S. (2002, forthcoming) *Economic Evaluations in Social Welfare: A Review of the Literature*. LSE Health and Social Care Discussion Paper
- MacDonald, G. (1996) 'Evaluating the effectiveness of social interventions', in A. Oakley and H. Roberts (eds) *Evaluating Social Interventions: A Report of Two Workshops Funded by the Economic and Social Research Council*. Ilford: Barnardos
- McKee, M., Britton, A., Black, N., McPherson, K., Sanderson, C. and Bain, C. (1999) 'Interpreting the evidence: choosing between randomised and non-randomised studies', *British Medical Journal*, Vol. 319, pp. 312–15
- Mannion, R. and Small, N. (1999) 'Postmodern health economics', *Health Care Analysis*, Vol. 7, pp. 255–72

- Martin, S. and Sanderson, I. (2000) 'Evaluating public policy experiments', *Evaluation: The International Journal of Theory, Research and Practice*, Vol. 5, No. 3, pp. 245–58
- Mays, N. and Pope, C. (1995a) 'Rigour and qualitative research', *British Medical Journal*, Vol. 311, pp. 109–12
- Mays, N. and Pope, C. (1995b) 'Observational methods in health care settings', *British Medical Journal*, Vol. 311, pp. 182–4
- Medical Research Council (2000) *A Framework for Development and Evaluation of RCTs for Complex Interventions to Improve Health*. London: Medical Research Council
- Mitton, L., Sutherland, H. and Weeks, M. (2000) *Microsimulation Modelling for Policy Analysis: Challenges and Innovations*. Cambridge University Department of Applied Economics Occasional Paper No. 65. New York: Cambridge University Press
- Morrell, C.J., Spiby, H. and Stewart, P., Walters, S. and Morgan, A. (2000a) 'Costs and benefits of community postnatal support workers: a randomised controlled trial', *Health Technology Assessment*, Vol. 4, No. 6
- Morrell, C.J., Spiby, H. and Stewart, P., Walters, S. and Morgan, A. (2000b) 'Costs and effectiveness of community postnatal support workers: randomised controlled trial', *British Medical Journal*, Vol. 321, pp. 593–8
- Moser, C.A. and Kalton, G (1993) *Survey Methods in Social Investigation*. Aldershot: Dartmouth Publishing Company
- Mugford, M. (2001) 'National evaluation of school breakfast clubs', presentation at workshop on 'Economic evaluation and social welfare: developing methods and learning from practical experience', University of Sheffield, 26 July
- National Centre for Social Research (NCSR) (2000) 'In-depth interviewing course', five-day training course run by NCSR, June
- Netten, A and Curtis, L (2000) *Unit Costs of Health and Social Care*. Canterbury: Personal Social Services Research Unit, University of Kent
- Newburn, T. (2001) 'What do we mean by evaluation?', *Children and Society*, Vol. 15, pp. 5–13
- NHS Centre for Reviews of Research and Dissemination (CRD) (2001) *Undertaking Systematic Reviews of Research on Effectiveness: CRD Guidelines for those Carrying out or Commissioning Reviews*. NHS CRD Report No. 4, 2nd edn. York: NHS CRD
- NHS Executive (1994) *Capital Investment Manual: Business Case Guide*. London: HMSO
- Nicolson, R., Fawcett, A., Moss, H., Nicholson, M. and Reason, R. (1999) 'Early reading intervention can be effective and cost-effective', *British Journal of Educational Psychology*, Vol. 69, pp. 47–62
- Nocon, A. and Qureshi, H. (1996) *Outcomes of Community Care for Users and Carers: A Social Services Perspective*. Buckingham: Open University Press

- Oakley, A. and Fullerton, D. (1996), 'The lamp-post of research: support or illumination', in A. Oakley and H. Roberts (eds) (1996) *Evaluating Social Interventions: A Report of Two Workshops Funded by the Economic and Social Research Council*. Ilford: Barnardos
- Patton, M. (1996) *Utilization-focused Evaluation*. London: Sage
- Payne, J. (2000) *Evaluating Training Programmes for the Long-term Unemployed*. PSI Report No. 871. London: Policy Studies Institute
- Pawson, R. and Tilley, N. (1997) *Realistic Evaluation*. London: Sage
- Pecora, P.J., Fraser, M., Nelson, K., McCroskey, J. and Meezan, W. (1995) *Evaluating Family-based Services*. New York: Walter de Gruyter
- Peled, E. and Spiro, S.E. (1998) 'Goal-focused evaluation', *Evaluation: The International Journal of Theory, Research and Practice*, Vol. 4, No. 4, pp. 455–68
- Pope, C. and Mays, N. (1995) 'Researching the parts other methods cannot reach: an introduction to qualitative methods in health and social services research', *British Medical Journal*, Vol. 311, pp. 42–5
- Pope, C., Ziebland, S. and Mays, N. (2000) 'Analysing qualitative data', *British Medical Journal*, Vol. 320, pp. 114–16
- Pratt, T. and Maahs, J. (1999) 'Are private prisons more cost-effective than public prisons? A meta-analysis of evaluation research studies', *Crime and Delinquency*, Vol. 45, No. 3, pp. 358–71
- Revolving Doors Agency (2000) *Mental Health, Multiple Needs, and the Police*. London: Revolving Doors Agency
- Robertson, A., Grimes, P. and Rogers, K. (2001) 'A short-run cost-benefit analysis of community-based interventions for juvenile offenders', *Crime and Delinquency*, Vol. 47, No. 2, pp. 265–85
- Ruhm, C. (2000) 'Parental leave and child health', *Journal of Health Economics*, Vol. 19, pp. 931–60
- Sassi, F. (2000) 'Methods for the evaluation of health care interventions', in R. Cookson, A. Maynard, D. McDaid, F. Sassi and T. Sheldon *Analysis of the Scientific and Technical Evaluation of Health Interventions*. Report to the European Commission, July. [http://www.lse.ac.uk/Depts/lsehsc/pdf\\_files/methods.pdf](http://www.lse.ac.uk/Depts/lsehsc/pdf_files/methods.pdf) (accessed in February 2002)
- Sassi, F., Archard, L. and Le Grand, J. (2001) 'Equity and the economic evaluation of health care', *Health Technology Assessment*, Vol. 5, No. 3, pp. 1–138
- Scott, S., Knapp, M., Henderson, J. and Maughan, B. (2001) 'Financial cost of social exclusion: follow-up study of antisocial children into adulthood', *British Medical Journal*, Vol. 323, pp. 1–5
- Sefton, T. (2000) *Getting Less for More: Economic Evaluation in the Social Welfare Field*. CASE Paper 44. London: Centre for the Analysis of Social Exclusion
- Shiell, A. and Hawe, P. (1996) 'Health promotion, community development and the tyranny of individualism', *Health Economics*, Vol. 5, pp. 241–47

Siegel, J., Torrance, G. and Russell, L. *et al.* (1997), 'Guidelines for pharmacoeconomic studies: recommendations from the panel on cost effectiveness in health and medicine', *PharmacoEconomics*, Vol. 11, No. 2, pp. 159–68

Stinnett, A., Murray, A., Mittleman, M., Weinstein, M., Kuntz, K., Cohen, D., Williams, L., Goldman, P., Staiger, D., Hunink, M., Tsevat, J., Tosteson, A. and Goldman, L. (1996) 'The cost-effectiveness of dietary and pharmacological therapy for cholesterol reduction in adults', in M. Gold, J. Siegel, L. Russell and M. Weinstein *Cost-effectiveness in Health and Medicine*. New York: Oxford University Press

Task Force on Principles for Economic Analysis of Health Care Technology (1995) 'Economic analysis of health care technology: a report on principles', *Annals of Internal Medicine*, Vol. 122, pp. 61–70

Thompson, R.W., Ruma, P., Schuchmann, M. and Burke, R. (1996) 'A cost-effectiveness evaluation of parent training', *Journal of Child and Family Studies*, Vol. 5, No. 4, pp. 415–29

Welsh, B.C. and Farrington, D.P (1999) 'A review of the costs and benefits of situational crime prevention', *British Journal of Criminology*, Vol. 39, No. 3, pp. 345–68

White, M. and Lakey, J. (1992) *The Restart Effect: Evaluation of a Labour Market Programme for Unemployed People*. London: Policy Studies Institute

Williams, A. (1992) 'Cost-effectiveness analysis: is it ethical?', *Journal of Medical Ethics*, Vol. 18, No. 1, pp. 7–11