Getting Less for More: Economic Evaluation in the Social Welfare Field

Tom Sefton

Contents

CASEpaper 44
November 2000

Centre for Analysis of Social Exclusion
London School of Economics
Houghton Street
London WC2A 2AE
CASE enquiries – tel: 020 7955 6679

Centre for Analysis of Social Exclusion

The ESRC Research Centre for Analysis of Social Exclusion (CASE) was established in October 1997 with funding from the Economic and Social Research Council. It is located within the Suntory and Toyota International Centres for Economics and Related Disciplines (STICERD) at the London School of Economics and Political Science, and benefits from support from STICERD. It is directed by Howard Glennerster, John Hills, Kathleen Kiernan, Julian Le Grand, Anne Power and Carol Propper.

Our Discussion Paper series is available free of charge. We also produce summaries of our research in CASEbriefs, and reports from various conferences and activities in CASEreports. To subscribe to the CASEpaper series, or for further information on the work of the Centre and our seminar series, please contact the Centre Administrator, Jane Dickson, on:

Telephone:	UK+20 7955 6679
Fax:	UK+20 7955 6951
Email:	j.dickson@lse.ac.uk
Web site:	http://sticerd.lse.ac.uk/Case

Editorial Note

Tom Sefton is a Research Officer at the Centre for Analysis of Social Exclusion and the Personal Social Services Research Unit at the London School of Economics.

Acknowledgments

This paper was prepared as part of on-going research funded by the Joseph Rowntree Foundation on developing the infrastructure for economic evaluation in the social welfare field. I am very grateful to the Foundation and to members of the Advisory Group set up to oversee the project, who participated in a helpful and vigorous discussion on an earlier draft and provided more detailed comments in writing. I would also like to thank the other members of the research team, John Hills, Martin Knapp, Howard Glennerster, Sarah Byford, Ann Netten, Jenni Beecham, Franco Sassi, and Dave McDaid, who provided many helpful contributions to the paper.

Abstract

Economic evaluation has an important role in helping to make decisions about the use of scarce resources in an explicit and rational manner, yet economic evaluation is not well-developed in many areas of social welfare. This paper looks at the reasons for this, focusing on what economists could do to redress the situation. It argues that standard approaches to economic evaluation may not always be appropriate, because of the nature of many social welfare interventions and because evaluators need to be able to address a broader set of evaluation questions. Economists could usefully contribute more to the debates that have concerned mainstream evaluators from other disciplines and modify their approach to evaluation accordingly. The paper concludes that in many areas of social welfare, economists should probably be *less* ambitious in terms of what they set out to achieve in terms of economic evaluation, but *more* ambitious in terms of the types of programme they can usefully help to evaluate and in terms of the range of techniques they are prepared to use, and give credence to, as part of an economic evaluation.

Introduction

This paper examines the current state of economic evaluation in the social welfare field, which would include services such as health care, employment, housing, regeneration, and social care. With the exception of the health care sector, rigorous economic evaluation is very rare in most areas of social welfare. The reasons for this are not well understood. Arguably, there is a lack of demand for economic input from those commissioning evaluation research and from non-economists involved in carrying out evaluation work. Where economists have been involved, this has often been at a late stage or at the margin, so they have been restricted in what they could do. Thus, in some cases, economists have not had the opportunity or adequate resources to apply their techniques properly. However, this paper starts with a presumption that part of the responsibility also rests with economists, who have often not tailored their approach to fit the requirements of evaluation in the social welfare field and generally have not worked closely with evaluators from other disciplines.

The paper begins by defining economic evaluation and describing standard approaches to it, including the association between economic evaluation and 'scientific' or experimental approaches to evaluation. It then examines some of the barriers to applying these standard approaches in the social welfare field, which are partly to do with the nature of many social policies and partly to do with incompatibilities with other evaluation perspectives. The last section discusses a possible way forward, including ways in which approaches to economic evaluation might be adapted to address some of the issues and concerns raised in the evaluation literature.

The paper deliberately adopts a critical perspective on economic evaluation in order to highlight those areas where current approaches might be developed further or where research carried out in other areas of economics could usefully be applied. This should not be seen as undermining the role of economic evaluation in the social welfare field. Economic evaluation has a vital contribution to make in helping to set priorities and make decisions in an explicit and rational manner about who receives certain services and the quantity and quality of services they should receive (e.g., H M Treasury, 1997; Drummond *et al*, 1997; Knapp, 1984; Stockdale *et al*, 1999; Byford, 2000). Economists sometimes seem to be alone in recognising that budgets are limited and that, as a result, difficult choices need to be made between alternative ways of

using scarce resources. Economists have also made a valuable contribution to evaluation in emphasising the importance of identifying explicit outcomes, considering the counterfactual (i.e. what would have happened in the absence of a particular intervention), and adopting a societal perspective or, in some cases, multiple perspectives. In many areas of social welfare, there is a strong case for more economic evaluation and for more 'scientific' approaches to evaluation.

Nor is it the case that many of the problems that are identified in the paper are unique to economic evaluation. Most of the problems are common to all forms of evaluation, although they are often addressed differently by other evaluators. Nevertheless, economists working in the evaluation field need to be aware of some of the difficulties with current approaches to economic evaluation, which may help to explain why they have not been more widely applied in the social welfare field. This may mean modifying techniques that have been developed in other policy areas and which may not be directly transferable to the social welfare field. In some areas, this process is already well under way, but in many areas there are still big gaps to be filled.

What is economic evaluation?

The purpose of economic evaluation is to inform decisions about the best use of limited resources. 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. Whilst all policy decisions are heavily influenced by political, ethical, and pragmatic considerations, economic evaluation has an important role to play in this process.

Spending money on a particular policy or programme means less money is available for other uses. A service may be effective to a degree in meeting the needs of its users, but it is inefficient if similar outcomes could be achieved at less cost, or if better outcomes could be achieved for the same cost.

Most types of evaluation focus mainly on the benefits. Economic evaluation takes into account both the costs and benefits of policies with a view to identifying the most cost-effective way of achieving policy objectives. What distinguishes economic evaluation from much evaluation work is that it explicitly recognises the resource constraints faced by decision-makers.

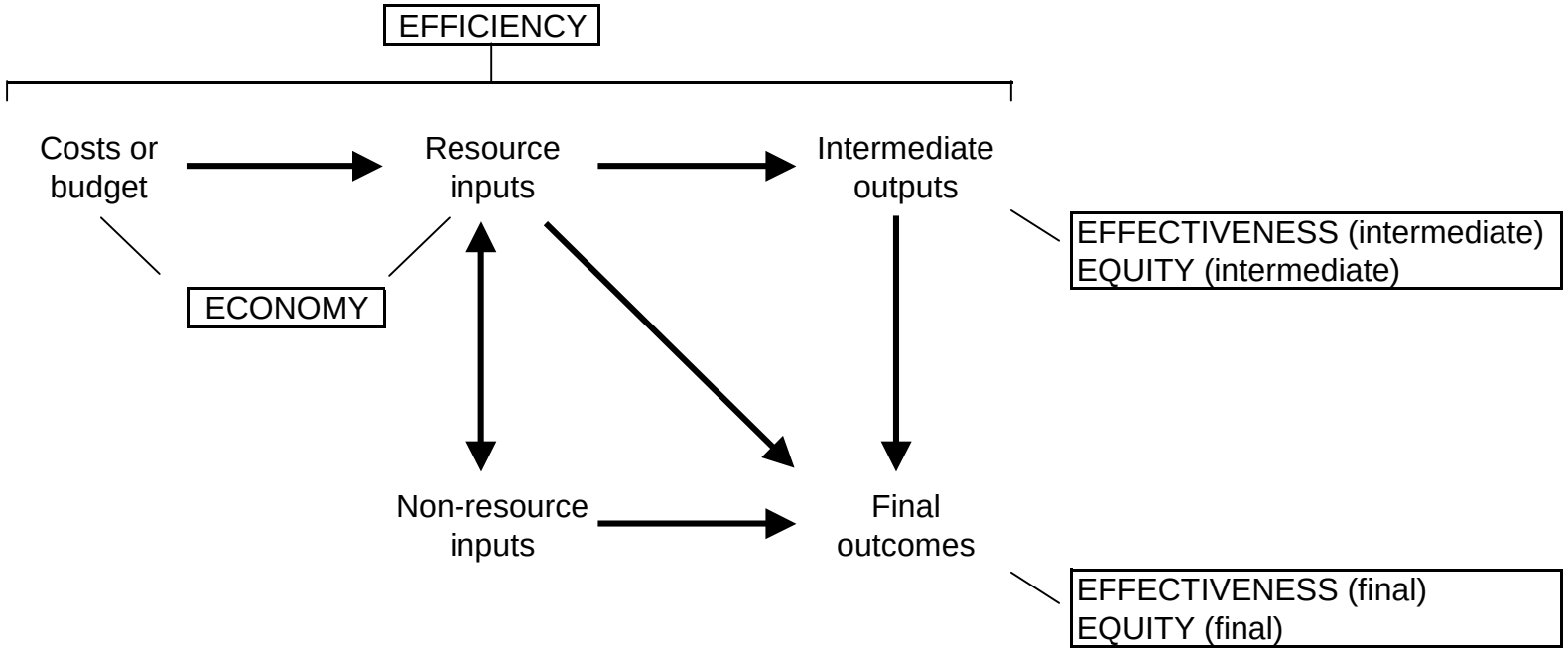
Economic evaluation usually involves a systematic attempt to identify and, where possible, measure, and compare the costs and outcomes of alternative policies. This is often based on a fairly basic framework, linking the inputs of the programme to outputs and outcomes. In the social care field, this framework has been developed further into the Production of Welfare approach (Knapp, 1984), which distinguishes between different types of inputs and outcomes (see Figure A).

One of the principles of economic evaluation is that it should involve a comparison between at least two options: for example a new 'pilot' scheme against an existing programme, or two variants of the same programme. The most 'complete' form of economic evaluation is cost-benefit analysis where the costs and benefits of alternative options are valued in monetary terms. Wherever possible, values are put on those goods or services that do not normally have a price attached to them. The option that has the greatest net benefit (or highest benefit / cost ratio) is preferred. Cost-benefit analysis has been employed in the evaluation of transport and development projects and, to a lesser extent, environmental and labour market policies.

More common is cost-effectiveness analysis where outcomes are measured in physical units, such as numbers of jobs of created, reductions in crimes committed, or improvements in a quality of life index. If two or more schemes have similar objectives, then it is possible to compare them in terms of their cost-effectiveness. For example, crime prevention schemes might be compared on the basis of cost per crime prevented. Where there are multiple outcomes, these are sometimes weighted to give an overall cost-effectiveness measure (or left unweighted as in cost-consequences analysis). Cost-effectiveness analysis is most developed in the health care sector, where it is used in the assessment of a wide range of treatments and policies, including pharmaceutical products, clinical procedures, and service arrangements.

This paper makes the distinction between evaluation, which is carried out after a policy or programme has been introduced, and appraisal, which is carried out beforehand. However, these same basic principles would apply to both.

Figure A: The Production of Welfare Framework



Source: Kendall and Knapp (2000)

Approaches to quantitative evaluation

Economic evaluation is an extension of, rather than an alternative to, impact evaluation (Rossi and Freeman, 1993). It is not possible to carry out a cost-benefit or a cost-effectiveness analysis if the effects of a policy or programme are unknown or cannot be accurately modelled. In order to conduct a standard economic evaluation, it is necessary to have quantitative data on the outcomes of specific interventions. This presents a number of challenges for the evaluator:

1) measurement: how to identify valid and reliable outcome measures;
2) attribution: how to isolate the effects of a particular intervention from other external factors that may also influence outcomes;
3) interpretation: how to generalise the results so that useful policy lessons can be drawn from the evaluation.

Quantitative evaluators, including economists, have tended to favour the use of experimental methods to isolate the effects of social welfare interventions from other factors that may also affect outcomes. Where experimental approaches are not feasible or are not comprehensive, modelling is often used to estimate or simulate the effects of a programme. These two approaches are described in more detail below.

Experimental approaches

Experimental approaches require at least two distinct treatment groups: one group of users is exposed to the programme that is being evaluated and the other group is excluded or is exposed to a different programme – often current 'best' practice. Outcomes for the 'experimental' and 'control' groups are compared and any differences attributed to the effect of the 'experimental' programme. Extraneous factors that would have affected outcomes for both groups are automatically stripped out. For example, a new Active Labour Market Programme that is introduced during a recession period may not appear to be very successful in improving employment rates among participants. But a comparison with the control group might show that those who did not take part experienced a significant deterioration in their employment prospects. Thus, the programme might be shown to have had a significant positive impact after all.

The validity of the experimental approach depends on there being a control group that is equivalent or very similar to the experimental

group, except for their exposure to the programme that is being evaluated. For this reason, researchers will often favour the use of randomised control trials, in which individuals or communities are randomly assigned to either the experimental or control groups. This avoids the danger that the results may be biased by other, possibly unobserved, differences between the underlying characteristics of the two groups. However, other 'quasi-experimental' techniques have been developed to attempt to control for known differences between the experimental and control groups, either by matching the two groups carefully or by using statistical analysis to adjust for differences in the characteristics of the groups.

Economic evaluation fits naturally with the experimental approach, since both are focused on the identification and measurement of outcomes that can be attributed to a particular programme. Economic evaluation goes a step further than impact assessment by measuring inputs, as well as outcomes, and by exploring the relationship between inputs and outcomes.

Modelling approaches

Whilst experimental techniques, and randomised control trials in particular, may be the preferred approach to evaluation, they are not always feasible in practice. Experiments need to be planned and conducted prospectively, whereas much evaluation work is carried out retrospectively, and so has to rely on what data is available after-the-event. Even when the evaluation strategy is thought out in advance, experimental approaches may be rejected for pragmatic or ethical reasons. For example, a full randomised control trial may be too expensive or it may be considered unethical to randomise when one intervention is believed to be less effective than the alternative. The heterogeneous nature of many social welfare interventions and multi-agency delivery can make it particularly difficult to control the interventions under investigation.

In these situations, modelling is another way of estimating the impact of particular policy interventions. Alternatively, modelling may be carried out on the back of a controlled experiment in order to extrapolate the results over a longer period or to generalise the results to other settings; this is fairly common practice in the health care field (Buxton *et al*, 1997). Models make use of secondary data and other sources, including expert opinion, to predict the outcomes of a particular intervention within a well-structured quantitative framework. For example, large-scale databases with information on individuals' medical

records and cholesterol levels have been used to model the long-term benefits of treatments designed to reduce patients' cholesterol levels (e.g., Stinnett *et al*, 1996). In the social welfare field, administrative records have been used in a similar way to model the effects of active labour market measures (e.g., Hasluck, McKnight, and Elias, 2000) and modelling is used extensively to estimate the impact of changes in the tax-benefit system (e.g., Evans, 1996; Blow and Crawford, 1997).

The advantage of models is that they are flexible; 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. But the outputs of a model are only as good as the quality of the inputs, so models are not a substitute for lack of data (although they can help to make the best use of limited data). They can also be somewhat obscure to those who were not involved in designing the model, including decision-makers, and they are open to bias in the way they are constructed and employed.

Key characteristics

The close association between economic evaluation and experimental approaches to evaluation means that they share many of the same characteristics:

1) method-driven: discussion of options is often rooted in the view that the best way to establish causality is by means of a randomised experiment (e.g., Rossi and Freeman, 1993). In practice, circumstances, ethical considerations, cost constraints, and time pressures necessitate compromises with this ideal. Thus, in the health care sector, there is a more or less explicit hierarchy of evidence, based on the evaluation methods used – with randomised control trials at the top, followed by quasi-experimental approaches, and expert opinion at the bottom (e.g., Drummond *et al*, 1997; Scott and Weston, 1998).
2) goal-based: the starting point and corner-stone for any economic evaluation is the identification of programme goals or objectives. This determines the choice of outcome measures against which the performance of a programme is then judged.
3) measurement: although in principle experimental designs could be conducted qualitatively, research cost considerations effectively rule out such approaches. The demand for precise estimates of net effects requires data that are quantifiable and uniformly collected, neither of which is conducive to qualitative analysis.

4) objectivity: is prized and evaluators are expected to maintain an independent position, so that they are not unduly influenced by the views of different stakeholders. For similar reasons, hard data is preferred to information on people's views or opinions, which is treated with a certain amount of suspicion, because of its subjective nature.
5) rigorous: the emphasis in most economic evaluation is on producing findings that are robust enough that another researcher using the same design in the same setting would achieve substantially the same results (Rossi and Freeman, 1993). Hence economic evaluators will tend to favour large-scale RCTs carried out in tightly controlled conditions. Less emphasis has tended to be given to whether the findings are representative of the programme as it will operate in practice and how the results might vary under different conditions, although economic evaluators are not alone in this.
6) ex post: this type of evaluation is most commonly undertaken after the programme has been in place for some time by which time it is hoped that teething problems will have been ironed out and the programme will have started to generate measurable outcomes.

Evaluation issues in the social welfare field

Evaluation, and quantitative evaluation in particular, is more straightforward if the programme is a distinct and well-defined intervention, with a large, short-term, and measurable outcome. Not surprisingly, few interventions meet all these criteria. Whilst evaluation is rarely a simple exercise, problems of measurement, attribution, and interpretation are more acute in some areas of social policy than others. Some specific problems are discussed below and may help to explain why quantitative evaluation is not more widespread.

1) multiple outcomes: social welfare interventions will typically have several outcome measures, which makes it more difficult to make comparisons between schemes – unless one performs better on all counts.
2) long-term outcomes: in some cases, the effects of a social policy intervention are not expected for many years, whereas decision-makers may be looking for early feedback.
3) qualitative outcomes: by their very nature, the outcomes of some social policies are not very amenable to measurement.

4) heterogeneity: many social welfare interventions are hard to define precisely, because of their fluid and multi-faceted nature.
5) local variations: these are another confounding factor, but are very much the norm in the social welfare field, because there is local autonomy in the delivery of many of these services. Increasingly, local bodies not only have responsibility for delivering services, but also for setting their own local objectives.
6) unit of analysis: often social policies are directed at areas or communities, rather than individuals. Random assignment is usually not possible with area-based initiatives and, where it has been tried (in the US), sample sizes are small. Even if a good initial match can be made between 'experimental' and 'control' areas, it is unlikely that circumstances would have remained similar throughout the evaluation period (Connell, 1995).
7) low level effects: the impact of social welfare interventions is often small (and, therefore, hard to detect) relative to the scale of the problem and compared to other external influences on outcomes.
8) user involvement: the active cooperation of users or clients is often crucial to the success of social welfare interventions. Thus, policies that are effective in one cultural setting or with one group of users may not work in other situations. This means that outcomes can be highly context-dependent, which makes it harder to produce results that are generalisable.

These difficulties are encountered in most evaluation work, whether it is in the social welfare field or outside it – and techniques have been developed to help address many of them. *However, in the case of social welfare programmes, it is more likely that an evaluation will have to deal with several of them within the same study, which compounds the problem.* For example, it may be possible to convert qualitative outcomes into a quantitative scale, which would capture significant changes in that outcome. But if, on top of this, the effects of a policy are expected to be relatively small and if the policy is interacting with other policies, then a quantitative scale may not be sufficiently sensitive to pick these up or it may be difficult to isolate the effect of that particular programme. A crude method for scoring the ease of evaluating different programmes – their 'evaluability' – is set out in Table 1, based on the programme's characteristics and the nature of the desired outcomes. Some of these obstacles can be overcome with sufficient resources, for example by extending the evaluation period to capture longer-term outcomes, but the resources available for evaluation are quite restricted in many areas of social welfare.

Table 1: Assessing the Evaluability of Social Welfare Programmes

Nature of outcomes	Well-defined objectives or not	Are the objectives of the programme clearly defined (i.e. Specific, Measurable, Achievable, Realistic, and Time-constrained)? (YES=1)
	Short or long-term	Can final outcomes be assessed within the time-frame of the evaluation? (YES=1)
	Unique or multiple	Is it reasonable to express the effects of the programme as a single (principal) outcome. (YES=1)
	Quantitative or qualitative	Are outcomes qualitative by nature or difficult to measure for other reasons (e.g. domestic violence)? (YES=0)
Nature of programme	Heterogeneity	Is programme a 'mixed bag' of different interventions or projects (or a fairly well-defined intervention)? (YES=0)
	Implementation	Is implementation likely to vary significantly within the programme? (YES=0)
	Unit of analysis	Is the programme directed at areas or communities and / or is sample size likely to be small for other reasons? (YES=0)
Context	Scale effect	Is the potential impact on outcomes likely to be small relative to the scale of the problem? (YES=0)
	External influences	Is the potential impact on outcomes likely to be small relative to other 'external' influences on outcomes? (YES=0)
	Active or passive clients	Is the effectiveness of the programme very sensitive to how clients choose to respond to it (or is something that is 'done' to essentially passive clients)? (YES=0)

Alternative perspectives on evaluation

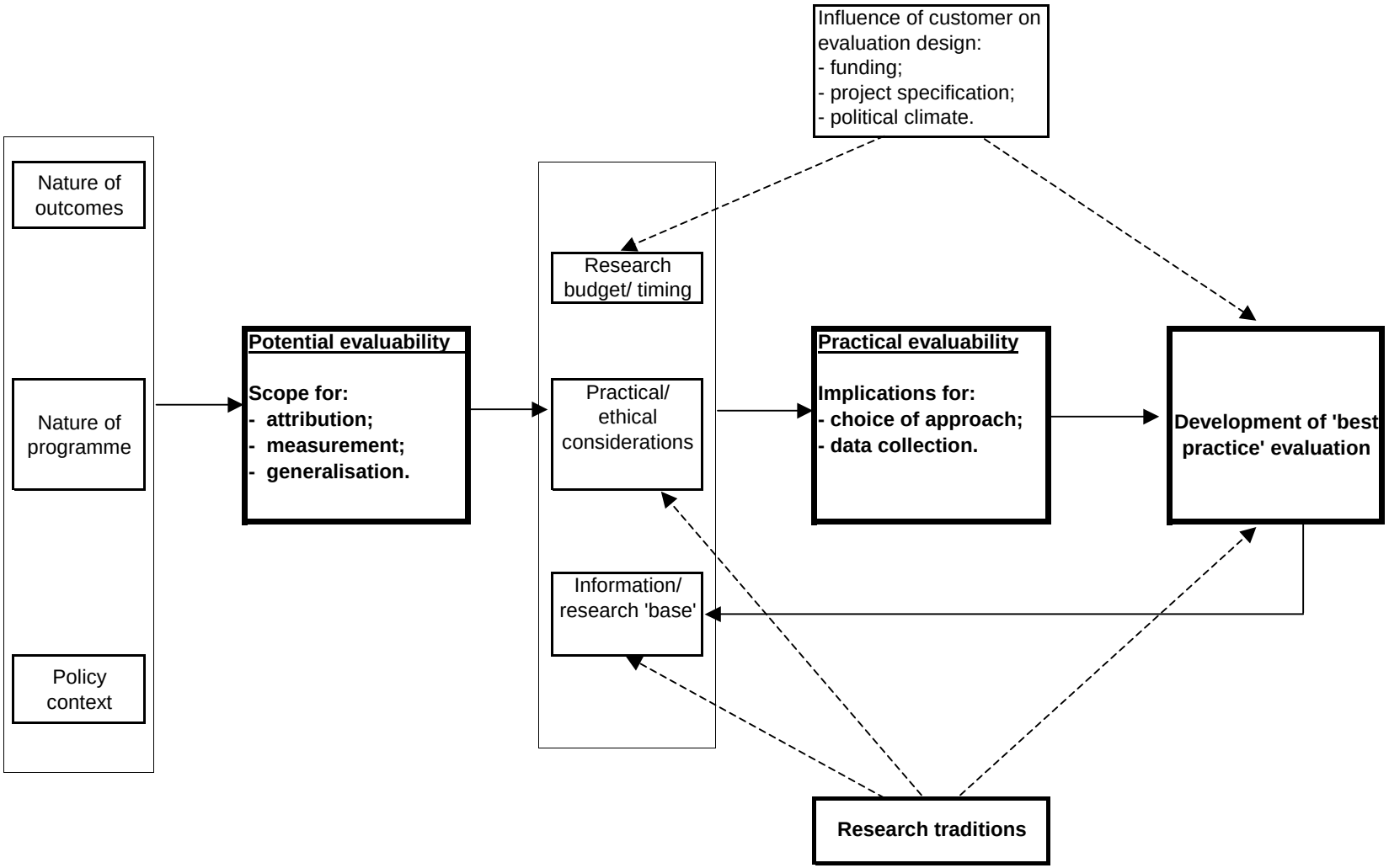
Evaluation is well-established as a discipline in its own right, distinct from, though with strong links to, the various social sciences from which it developed. Economic evaluation needs, therefore, to be set in the context of evaluation research more generally. One of the reasons economic evaluation has been slow to establish itself in some areas of welfare policy is that many social welfare interventions are not conducive to standard economic approaches, as discussed above, and economists have not always been very adaptable to these varying requirements.

Another reason is that most economic evaluation has focused on assessing the value-for-money of individual programmes, whereas policy-makers and other customers of evaluation research are interested in a much wider set of evaluation questions, including a better understanding of the implementation process or how to build a learning process into the development of new policies. Again, economists do not appear to have been very effective in incorporating an economic component into these other types of evaluation.

Last but not least, there are philosophical differences of opinion between evaluators, in particular the long-standing debate between the 'scientific' or 'positivist' approaches at one end of the spectrum and constructivism at the other. More recently, there have been attempts to establish 'synthesis' approaches, which lie between these two extremes. Examples include "realistic evaluation" (Pawson and Tilley, 1997), "theory-based" evaluation (Chen, 1990), and "utilization-focused" evaluation (Patton, 1996). Each approach has its own conceptual framework, but most advocate methodological pluralism – a mix of quantitative and qualitative techniques – to suit the particular requirements of the evaluation. Economists have tended to remain firmly in the 'scientific' camp and have not been closely involved in this wider debate.

Figure B shows the various factors that are likely to influence the development of economic evaluation in different areas of social welfare. We have already discussed the "potential evaluability" of programmes as determined by the programme's characteristics. What follows is a brief discussion of the issues that have pre-occupied evaluation researchers – the "research traditions" - and some of the implications for how we think about the economic evaluation of social welfare interventions.

Figure B: Scope for Evaluation of Social Welfare Interventions



Theory-based evaluation

One of the criticisms of 'scientific' approaches to evaluation is they have become largely atheoretical (Chen, 1990). The focus of many studies is on the overall relationship between inputs and outcomes with little concern for how these outcomes are generated – the 'black box'. This kind of analysis may show that a programme is not working, but it would not help to determine what the problem is. For example, is it that the programme is not being implemented properly or is that the theories or assumptions underlying the programme are 'wrong'?

Other commentators have stressed the context-dependence of most social policy interventions. The question we should be asking is not whether a policy works, but rather what works, for whom, and in what circumstances. Experimental evaluation, which tends to focus solely on the net effects of a programme, will often produce inconsistent results – and this is to be expected because people will respond differently to the same intervention depending on their circumstances. Without an understanding of *why* programme impacts vary between places or over time, the policy-maker is left none the wiser (Pawson and Tilley, 1997).

This critique could also be levelled at much economic evaluation work, even though econometric modelling provides the necessary statistical tools to examine variations in costs and outcomes between individuals. However, there are examples of economic evaluations that do focus on inter-individual differences (e.g., Knapp *et al*, 1992) and, in some cases, the reasons for those differences (e.g., White and Lakey, 1992). There is also plenty of research by economists on what it is that makes policies work, including a whole literature on incentive design, but this has not been integrated into mainstream economic evaluation.

Various approaches have been put forward in the evaluation literature, which are designed to give more weight to the theoretical basis for evaluation. Programme theory is a set of explicit and testable assumptions or hypotheses about how a programme is supposed to achieve its goals. In "Realistic Evaluation", these hypotheses take the form of Context-Mechanism-Outcome configurations (i.e. a specified mechanism will produce given outcomes in certain contexts). A programme may be based on a number of assumed CMOs, which can be tested in practice using whatever methods are most appropriate. An illustrative example is given in Table 2, which shows how an evaluation can be much more informative if, as well as monitoring outcomes, it also seeks to test which mechanisms are operating in practice and explores the context(s) in which they are likely to be triggered. The purpose of 'Realistic Evaluation' is to identify effective CMOs on the basis that this

information is more transferable – and therefore more useful to decision-makers - than one-off results from experimental evaluations.

Theory of Change approaches also focus on how and why an initiative works as a central part of the evaluation. It was developed as a tool for evaluating community-wide initiatives (Connell, 1995), such as early intervention schemes for children, but has potential applications to other areas of social welfare (e.g., Hughes and Traynor, 2000). The Theory of Change model takes for granted that programmes are based on implicit or explicit theories about how and why the programme is supposed to work. The first stage of an evaluation is to work with programme managers to define:

- what the programme wants to achieve and by when;
- how they are going to get there;
- whilst being prepared to adapt their methods as they go along.

The Theory of Change model fits within an input-outcome framework, which is consistent with standard economic approaches to evaluation. The evaluator starts by defining long-term objectives and works backwards from the endpoint through the steps required to get there. Early stage or intermediate objectives are then established for each step, so that the programme can be evaluated, and if necessary modified, at any stage. Whereas long-term objectives are more likely to focus on outcome measures, shorter term objectives are more likely to focus on process measures. Economists are often critical of process measures, because they do not directly measure the achievement of objectives. But, within a theoretical framework, such as the Theory of Change, process measures may be the only effective way of assessing whether a programme's objectives are likely to be achieved.

In practice, the appropriate emphasis to be given to programme theory depends on the nature of the programme and the decision-makers' own requirements. Theory-based approaches to evaluation are more relevant to new policies and to policies that are likely to be context-dependent than to established policies that work in a more mechanical fashion. Julnes et al argue that programme theory is more like a set of Russian stacking dolls than a black box – and that the intended use of the evaluation should determine how far the evaluator goes in seeking to unpack and understand how a particular programme is operating (Julnes *et al*, 1999).

Table 2: Illustrative Example of Realistic Evaluation

Use of CCTV to Reduce Car-Related Crime

Possible Mechanisms	**Contexts**	**Expected outcomes**
'you've been framed' mechanism: CCTV could reduce crime by deterring potential offenders who will not want to risk apprehension, and conviction by evidence captured on videotape.	*'criminal clustering' context:* a given rate of car crime may result from a few busy criminals or many 'minor' offenders. *'lie of the land' context:* cars parked in CCTV blind spots will be more vulnerable if the mechanism is increased chances of apprehension through evidence on videotape.	• if 'busy' car thieves are apprehended, this could significantly reduce crime both in the car park where CCTV was fitted and elsewhere. • could reduce crime in car parks with CCTV, though may be some displacement within car parks (to blind spots) and to car parks without CCTV.
'effective deployment' mechanism: CCTV may facilitate deployment of security staff or police officers towards areas where suspicious behaviour is occurring.	*'resources' context:* effective deployment will be more feasible in city car parks with greater security presence than in more isolated or dispersed car parks.	• should lead to less car crime on the car parks where CCTV is installed, though car crime may be displaced elsewhere.
'publicity' mechanism: CCTV could symbolize efforts to take crime seriously. Potential offenders may be deterred by the greater risk they believe to be associated with committing car crimes in car parks.	*'surveillance culture' context:* as the use of CCTV spreads through all walks of life, the efficacy of the publicity given to CCTV in car parks may be enhanced or muted by the overall reputation of this form of surveillance.	• may lead to a general reduction in car crime if CCTV is seen as part of wider clampdown on car crime.

Possible Mechanisms	Contexts	Outcomes
'*memory jogging*' *mechanism:* CCTV and notices indicating that it is in operation may remind drivers that their cars are vulnerable, and they may be prompted to take greater care to lock them and operate any security devices.	*'state of the art' context:* if security devices are fiddly and need to be activated, then drivers may be less likely to use them habitually and so may be more responsive to promptings, such as CCTV signs.	• should lead to less car crime on CCTV car parks, provided additional safety precautions are effective in deterring car thieves.
'*appeal to cautious*' *mechanism:* cautious drivers, who are sensitive to the possibility that their cars may be vulnerable and are habitual users of various security devices, may fill car parks with CCTV.	*'style of usage' context:* if the dominant mechanism is increased confidence, CCTV may have little impact in car parks where usage is dictated by other factors (e.g. location to nearby facility and where users have few alternatives).	• may lead to reduced car crime on CCTV car parks, but car crime could be displaced elsewhere (i.e. to where the more vulnerable cars are now parked)

Source: adapted from Pawson and Tilley (1997), pp 78-80.

Treatment of goals

Most evaluation, including economic evaluation, is goal-based. It is assumed that all programmes should have clearly defined objectives which are set out in advance and against which their performance can be assessed. In practice, a programme's objectives may not be so well-defined and may vary between stakeholders. Even where there are 'declared' goals, these may be different to those reflected in the actual operation of the programme – the 'operative' goals.

"Goal-focused evaluation" recognises that goals are rarely as clear-cut as many evaluators would like them to be. The clarification of goals 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 goals are identified, compared with operative goals, discussed with stakeholders, and then revised if appropriate. Only then does evaluation proceed to assessment of whether these goals are being achieved. Economists, on the other hand, have tended to see ill-defined goals as a reflection of poor policy design. Defining goals or objectives is seen as the responsibility of policy-makers and not as part of the evaluation process. None of this denies the importance of having clearly defined goals as the 'corner stone' for evaluation, but it does have implications for the way goals are treated by the evaluator.

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. For these kinds of programmes, the evaluation framework needs to be more flexible than standard approaches to evaluation to allow for the fact that goals may vary between projects and, perhaps, over time.

A separate issue is that evaluation typically focuses on individual programmes. Interactions with other policy initiatives are controlled for, rather than explored. "Target-oriented evaluation", on the other hand, recommends that evaluators begin with a set of more broadly defined policy targets and then assess which policies or combinations of policies are best suited to achieving these goals (Schmid, 1997). Thus, the effectiveness of individual programmes is assessed against wider policy aims, as well as more programme-specific objectives.

Prospective evaluation

The UK Treasury has in the past adopted a fairly narrow definition of evaluation: "checking afterwards whether objectives have been fulfilled" (H M Treasury, 1988). But the kind of rational ex post analyses once favoured are no longer adequate. The style of policy formulation has changed noticeably under the current Government and this is putting different demands on evaluators (Martin and Sanderson, 2000). In particular, the increasing use of pilot programmes, and the way in which they are being employed, means that evaluators need to be able to assess and learn from programmes prospectively, as well as evaluate them retrospectively. Policy-makers are often committed to particular policies, so evaluation of pilots is not so much about experimentation (i.e. whether a programme can be shown to have worked) as about 'exemplification' (i.e. providing examples of good practice or 'trailblazers' for others to follow).

This has several implications for the evaluation of programmes, including greater emphasis on refining new programmes, demonstrating their benefits, and disseminating good practice. This alters the role of the evaluator, who may need to be more actively involved in the development of new policy. This role may not be easy to combine with a more traditional 'desk-based' evaluation, which is still the basis for most economic evaluation. The timing of evaluation work is also affected. Policy-makers are looking for relatively quick results, even though the objectives may be very long-term.

What is needed is a more iterative approach to evaluation, whereby programmes are continually monitored, evaluated, and modified over time. One of the advantages of the Theory of Change approach discussed above is that evaluation can be carried out at any point in terms of the achievement of intermediate objectives and the probability of achieving longer term outcomes.

Alongside this development, it is now generally accepted that an important part of any evaluation is checking that a programme is being implemented in the way it was intended to be – what is sometimes referred to as process evaluation or treatment / programme fidelity. This recognises that the way a programme is implemented will often be different to the way it has been implemented elsewhere or the way it is described in published policy documents and that this can have a significant impact on outcomes. It is not simply a case of waiting for teething problems to be ironed out.

Limits of positivism

A more fundamental critique of scientific approaches is their assumption that there is an objective truth out there waiting to be discovered. Constructivists, on the other hand, would argue that there are as many 'realities' as there are 'actors'. What we see depends on who we are and how we look – and, therefore, evaluation can never be value-free or objective (Kushner, 1996). This has far-reaching implications for the way evaluation is approached, including greater participation of different stakeholders in the evaluation process. The role of the evaluator is to seek an agreed or negotiated consensus between stakeholders, rather than attempt to get at the 'truth' through scientific methods.

Taken to an extreme, constructivist evaluation could become little more than a description of different people's experiences of a programme. Less extreme approaches attempt to synthesise the views of different stakeholders, balancing or negotiating between competing views. Often, the process of involving stakeholders in the evaluation can be as, if not more, useful than the findings themselves (Patton, 1996).

Although true constructivists would reject any compromise between the two opposing views of the world (Guba and Lincoln, 1989), there are ways in which scientific approaches to evaluation could and, in some cases, already have been modified to accommodate the constructivist critique, including:

- giving less weight to measurement and precision and more weight to qualitative data;
- being less pre-occupied with objectivity and engaging more closely with stakeholders;
- conducting "sensitivity checks": varying assumptions to reflect different perspectives or value positions;
- allowing program theory to emerge during the evaluation, rather than seeking to 'impose' a priori theories.

Evaluation criteria

A specific criticism of economic evaluation, including by many economists, is that it tends to focus on the efficiency of programmes and ignores other important values, such as "equity" or "humanity". There is a danger in making efficiency the overriding criterion and seeking to make economic problems into purely technical ones.

This is sometimes justified on the basis that equity issues are best dealt with by means of government (lump-sum) transfers and that programme evaluation should, therefore, concentrate on efficiency. A programme is worthwhile if the benefits outweigh the costs (i.e. if the

winners could, in principle, compensate the losers and still have some left over). In practice, compensation is never paid in full to the losers, so equity is always likely to be an issue (Fraser, 2000). It is unfair to say that economists have ignored the distributional effects of public policies, since many studies focus specifically on this issue. But, it is the case that distributional and other equity issues are not an *integral* part of standard economic analyses in practice, despite strong recommendations in principle (e.g. Gold *et al*, 1996).

In the case of cost-benefit analysis, inequity is, arguably, built into the way economists attach monetary values to benefits and costs, using 'Willingness To Pay' (WTP) as a measure of worth. WTP is partly a matter of 'ability-to-pay', so it effectively gives more weight to the preferences of the better-off. Whilst economists would not claim that valuation techniques are 'value-free', these values are sometimes buried within the analytical framework (House, 2000).

The 'humanity' of programmes is a generic term, which encompasses a range of other values, such as fairness, justice, and empowerment – none of which are covered by the usual criteria. One of the reasons for being interested in process is that the way a programme is carried out can be significant in itself, quite independently of how this affects the achievement of programme goals. For example, the same goals could be pursued in ways that either empower or disempower those involved, depending on how prescriptive they are and whether they are more or less 'user-focused'. Empowerment and other values ought to be included within the evaluation framework, either as goals in their own right 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 happening (e.g. Kendall and Knapp, 2000). But this does a present a challenge for economists and will, in many cases, require more attention to be given to processes, as well as outcomes.

Way forward

There should be little doubt that 'policy' decisions about social welfare programmes are in part economic ones and that there is, therefore, an important role for economic evaluation. Some commentators have detected a shift away from effectiveness towards efficiency as the main evaluation criterion (e.g. Shaw, 1997). However, many areas of social

policy are still relatively untouched by economic thinking. Many of the reasons for this have already been discussed.

Knapp suggests that the state of economic evaluation in a particular area will typically go through five historical phases, based on his experience of applying economics to different areas of social and health care (Knapp, 1999):

1) blissful ignorance: little concern for cost or value-for-money. Assumption is that budgetary growth will solve society's problems;
2) unbridled criticism: reaction against cost constraints imposed by economic realities. View is that decisions should be made on the basis of need and/or professional opinion, rather than efficiency considerations;
3) undiscriminating utilisation: recognition that economic evaluation has a role to play in resource allocation decisions, but techniques are under-developed: terms are used inconsistently and design flaws pervade;
4) constructive development: techniques become more sophisticated and are adapted to increase their relevance. Economic studies begin to inform, though not dominate, decision-making by policy-makers and others;
5) sublime sophistication: economic methodologies are widely used, conducted well, and interpreted appropriately.

In parts of the social care field, he argues that we have reached the fourth stage in the UK and the same could be said of other specific areas of social welfare, such as the evaluation of welfare-to-work programmes. However, in many areas of social welfare, we are still at the second or third stage. One of the features of these earlier stages is that even where evaluation studies include an economic component, it is often tacked on to the end of an existing study and carried out independently of other parts of the evaluation. Often, this division between the economic and non-economic components of evaluation persists even when economic evaluation has broken through to the fourth stage.

There are basically two approaches to extending the use of economic evaluation in those areas where it has not yet developed beyond the second or third stage:

a) promote the application of 'best practice' economic evaluation techniques, persuading sceptics of the necessity and usefulness of this kind of analysis as a complement to, or an extension of, existing approaches to evaluation. This school of thought would probably favour technical solutions to evaluation problems,

'borrowing' from other policy areas where economic evaluation is more advanced, such as health economics.

b) recognise that current approaches may not be appropriate in many social policy contexts and therefore seek to adapt or modify standard economic approaches to economic evaluation to address some of the issues discussed above, whilst maintaining the same basic principles of economic analysis;

There is a place for both approaches, depending on the circumstances. Certainly, there are some areas where existing techniques are applicable and what is needed is economists who are willing to work in these areas and clients who are willing to fund their work. However, this paper has argued that this is not the whole story and that there are good reasons why standard approaches to economic evaluation may need to be adapted to meet the challenges faced by evaluators in the social welfare field, whether they be economists or not. In summary, these reasons include:

1) flexibility: policies are more developmental than allowed for by standard approaches to economic evaluation. The evaluation framework needs to allow for the fact that goals may differ from those set out in policy documents and that they may even change over time;
2) theoretical basis: much 'run-of-the-mill' evaluation has become too method-driven, so much so that it is possible to carry out an evaluation without an understanding of how the programme operates. A greater emphasis on theory-based evaluation – and getting inside the 'black box' – would help evaluators to understand why policies work or fail in different circumstances. This should provide more transferable lessons for decision-makers than estimates of net effects and would also provide a sounder basis for developing intermediate outcome measures (by providing testable theories that link intermediate to final outcome measures);
3) different perspectives: economists need to engage more closely with the various stake-holders who are either funding, administering, or at the receiving end of social welfare interventions. This is not just about ensuring they are not excluded from the analysis of costs and outcomes, although that is important. It is also about involving them in defining programme goals, using their knowledge to develop programme theories as to how these goals may be achieved, examining incentive structures

within the operation of the programme, and helping to design the evaluation strategy;

4) additional criteria: there is a need to incorporate additional criteria into the evaluation framework, including equity as well as other criteria, which are always likely to be important in a social welfare context;

5) methodological pluralism: a strict hierarchy of techniques is probably not the most helpful way to approach evaluation in many areas of social policy. Most mainstream evaluators recognise that they need to be able to employ a range of techniques, including qualitative and quantitative methods, to match the requirements of different evaluations.

Some of these challenges are more straightforward than others for economists to take on board. Economists have always put a strong emphasis on developing the theoretical underpinnings of their research, so this critique ought to be easier to address. Where most progress has been made and where economists have most to offer, is where decisions are more likely to be dictated by economic, as opposed to sociological or psychological, factors. For example, there is a big literature on labour market incentives which has informed the evaluation of welfare-to-work programmes. In other areas where economics may have less to contribute to 'theory-generation', economists still need to be aware of the theories and contexts that underlie a social programme, since these have implications for economic evaluation, including, for example, the sources of cost and outcome variations that may need exploring.

In some cases, the challenges will be harder for economists to face, because there is a potential conflict with standard approaches to economic evaluation. Economists have become very reliant on quantitative methods of evaluation, so they are often reluctant to incorporate qualitative evidence into their research. Some economists would probably argue that a piece of analysis is not an economic evaluation if it does not include fully quantified, and preferably monetised, estimates of costs and outcomes. However, in some areas of social welfare, this is just not possible, at least initially, given the nature of outcomes and the availability of data.

In other cases, economists have already developed their own way of addressing the issue, which is different to the approach taken by other disciplines; there is, therefore, a need for greater understanding, and scope for beneficial cooperation, between disciplines. For example, it is possible to argue that welfare economics is the ultimate stakeholder approach, since it is based on the preferences of those who use a

particular service or are in any other way affected by it, either positively or negatively. Stated preference techniques, where people are asked to put a monetary value on defined outcomes, clearly do involve economists engaging directly with a key stakeholder – the general public. Indeed, some economists would argue that these are the only views that should really count, since it is they who receive the benefits and ultimately bear the cost of social welfare interventions. This is not the place to discuss the controversies surrounding monetary valuation, but, in any case, most economic evaluation in the UK is largely desk-based and rarely involves the use of stated preference techniques. Having said this, the emphasis given by economists to taking a societal perspective and, therefore, to a *comprehensive* assessment of costs and outcomes, has in some cases led to the interests of important stakeholders being taken into account, where they had previously been overlooked by evaluators. In the social care and long-term health care field, for example, economists have done quite a lot of work on the experiences of unpaid carers to ensure that evaluation incorporates the impact of policy on this group (e.g., McDaid and Murray, forthcoming).

Whereas economists are primarily interested in the value or weight people attach to different outcomes, other social researchers would seek to involve stakeholders in a much broader set of issues, including:

1) helping to clarify the objectives that are driving the operation of the programme in practice;
2) identifying the assumptions or theories upon which the design of the programme is based;
3) checking whether the programme is being implemented as intended;
4) helping to understand why a programme might be having certain, perhaps unanticipated, effects;
5) exploring how people respond to a particular intervention, what they feel about it, and what improvements they would recommend;
6) assessing what impact those involved would attribute to a particular intervention;
7) encouraging participants to be more closely involved in, contribute to, and learn from, the evaluation process.

Thus, stakeholders can be involved in ways that most economists would not normally consider. This input is likely to be particularly valuable in certain situations; for example, evaluators are more dependent on stakeholders to identify the impact attributable to a

particular intervention if the use of a control group is not feasible for whatever reason.

Evaluation hierarchy

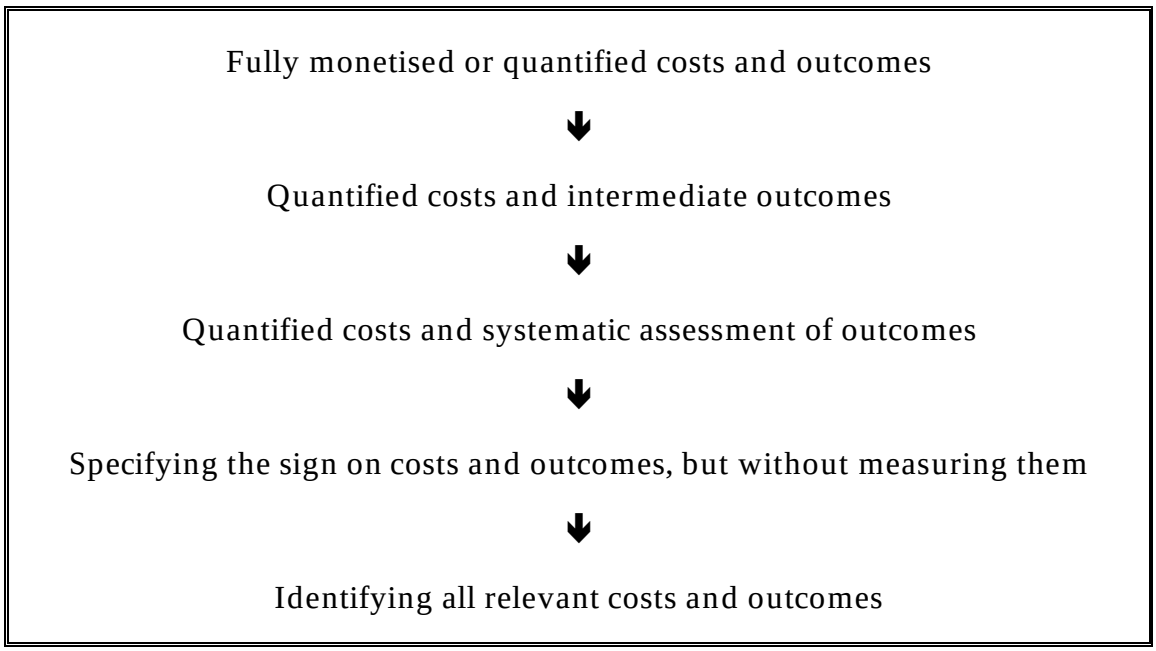
So, what can economists do if, as will often be the case, the ideal is not achievable? How, in practical terms, might standard approaches to economic evaluation be adapted? It is helpful first to consider what the minimum requirements are for economic evaluation. At its most basic level, these would include:

1) well-defined objectives;
2) assessment of costs and relevant outcomes;
3) a counterfactual (e.g. a suitable comparator)
4) an evaluation procedure (e.g. cost-effectiveness analysis).

On this basis, it is possible to develop a kind of hierarchy of approaches (see Table 3). The top level is what most economists would refer to as economic evaluation and covers cost-benefit, cost-effectiveness, and cost-consequences analysis. Ideally, you would want to be at this top level or, at least, moving in that direction. However, costs and final outcomes may be hard to measure with precision at reasonable cost and in the time available. For example, experimental approaches may not be able to provide information on key outcomes or stated preference techniques may involve too much uncertainty to allow statistical significance to be established at reasonable cost. Therefore, the evaluator may be forced to look at intermediate outcomes and to define and, if possible 'test', the theories that link intermediate and final outcomes.

Where quantification of any kind would be very hard, it may still be useful to carry out a systematic assessment of costs and outcomes, including qualitative outcomes. Economists can contribute by helping to develop the analytical framework for assessing outcomes. Even if outcomes cannot be quantified, anything that allows greater comparability between the outcomes of different interventions is a step in the right direction.

Table 3: Hierarchy for Economic Evaluation



Alternatively, it may be possible to put a likely sign on the costs and outcomes of a change in policy. In principle, this can be done without even trying to assess costs and outcomes directly. Theory might suggest that positive outcomes will result provided certain conditions are met. For example, an early evaluation of GP fundholding argued that the scheme's success depended on GPs being willing to use their purchasing power to improve services for their patients, being able to keep within budget, and not taking advantage by cream-skimming the best (i.e. cheapest) patients (Glennerster *et al*, 1994). Their analysis tested whether these assumptions or theories held in practice, providing a good indication as to whether or not the programme was likely to generate the positive outcomes expected by policy-makers - and, thus, would fit within this hierarchy of economic evaluation.

Finally, it may still be useful for an economist to encourage people involved in a programme to think through and identify what the possible costs and outcomes are and to start monitoring these, even if an economic evaluation cannot be conducted at this stage. Economists need to be involved as early as possible in an evaluation, in the process of defining objectives that provide a sound basis for subsequent evaluation, in establishing a framework for considering costs and outcomes systematically, in developing valid outcome measures that can be monitored over time and used to evaluate the programme at some future date, and in thinking about what comparators could be used as a benchmark for the programme to be evaluated.

As we move down the hierarchy in Table 3, it is less likely that the evaluation techniques normally favoured by economists will be applicable. Fewer outcomes will be quantifiable, so more qualitative research methods will often be more useful (though the results could still be fitted within a more systematic evaluation framework). It is also less likely that the use of control groups will be feasible, so other ways may need to be found to demonstrate that a programme has had a significant impact, including the combined use of several methods or data sources (i.e. triangulation) and a greater reliance on defining and testing programme theories. As part of this process, economists may need to engage with stakeholders for a wider variety of purposes.

Thus, in many areas of social welfare, economists probably need to be less ambitious in terms of what they set out to achieve and more open to less demanding forms of economic evaluation. At the same time, economists need to be more ambitious in terms of the types of programme they can usefully help to evaluate and *more* eclectic in terms of the range of techniques they are prepared to use, and give credence to, as part of an economic evaluation.

References

Blow, Laura and Crawford, Ian (1997), *The distributional effects of taxes on private motoring*, London: Institute of Fiscal Studies.

Buxton, Martin J. *et al* (1997), 'Modelling in economic evaluation: An unavoidable fact of life', *Health Economics* 6: 217-227.

Byford, S. (2000), "The Application of health economic techniques to social welfare", unpublished manuscript.

Chen, Huey-Tsyh (1990), *Theory-driven Evaluations*, Beverly Hills, CA: Sage.

Connell, James P. (ed.) (1995), *New Approaches to Evaluating Community Initiatives: concepts, methods, and contexts*, Queenstown, Md: Aspen Institute.

Drummond, Michael F. *et al* (1997*), Methods for the Economic Evaluation of Health Care Programmes*, Oxford; New York: Oxford University Press.

Evans, Martin (1996), *Giving credit where it's due: The success of Family Credit reassessed*, Welfare State Programme Discussion Paper WSP / 121, London: STICERD, London School of Economics.

Fraser, Neil (2000), 'Economic Policy Analysis', in Hills, J. and Fraser, N, (eds), *Public Policy for the 21st Century: Essays in memory of Henry Neuburger*, Bristol: Policy Press.

Glennerster, Howard *et al* (1994), *Implementing GP Fundolding: Wild Card or Winning Hand?*, London; Bristol Pa: Open University Press.

Gold, Marthe R. (1996), *Cost-effectiveness in health and medicine*, New York: Oxford University Press.

Guba, Egon G. and Lincoln, Yvonna S. (1989), *Fourth Generation Evaluation*, London: Sage.

Hasluck, Chris, McKnight, Abigail and Elias, Peter (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, Leeds: The Stationary Office.

H M Treasury (1988), *Policy Evaluation: A Guide for Managers*, London: HMSO.

H M Treasury (1997), *Appraisal and Evaluation in Central Government*, London: HMSO.

House, Ernest, R. (2000), 'The Limits of Cost-Benefit Analysis', *Evaluation: The International Journal of Theory, Research and Practice* 6(1): 79-86.

Hughes, Mike and Traynor, Tish (2000), 'Reconciling Process and Outcome in Evaluating Community Initiatives', *Evaluation: The International Journal of Theory, Research and Practice* 6(1): 37-49.

Julnes, George, Mark, Melvin M, and Henry, Gary T. (1999), 'Promoting Realism in Evaluation', *Evaluation: The International Journal of Theory, Research and Practice* 4(4): 483-504.

Kendall, Jeremy and Knapp, Martin (2000), 'Measuring the Performance of Voluntary Organisations', *Public Management* 2:1 105-32.

Knapp, Martin (1984), *The Economics of Social Care*, Basingstoke and London: Macmillan.

Knapp, Martin (1999), 'Economic Evaluation and Mental Health: Sparse Past…Fertile Future?', *Journal of Mental Health Policy and Economics 2(4).*

Knapp, Martin *et al* (1992), *Care in the Community: Challenge and Demonstration*, Aldershot: Ashgate.

Kushner, Saville (1996), 'Limits of Constructivism in Evaluation', *Evaluation: The International Journal of Theory, Research and Practice* 2(2): 189-200.

Martin, Steve and Sanderson, Ian (2000), 'Evaluating Public Policy Experiments', *Evaluation: The International Journal of Theory, Research and Practice* 5(3): 245-58.

McDaid, David and Murray, Joanna (forthcoming), 'Carer Burden: The Difficulties and Rewards of Caregiving', in Warner, M. *et al*, *Alzheimer's Disease Policy and Practice*, Oxford: Radcliffe Medical Press.

Patton, Michael Q. (1996), *Utilisation-focused Evaluation*, Beverly Hills, CA: Sage.

Pawson, Ray and Tilley, Nick (1997), *Realistic Evaluation*, London: Sage.

Peled, Einat and Spiro, Shimon E. (1998), 'Goal-focused Evaluation', *Evaluation: The International Journal of Theory, Research and Practice* 4(4): 455-468.

Rossi, Peter H. and Freeman, Howard E. (1993), *Evaluation: A Systematic Approach*, Beverly Hills, CA: Sage.

Schmid, Gunther (1997), 'The Evaluation of Labour Market Policy: Notes on the State of the Art', *Evaluation: The International Journal of Theory, Research and Practice* 3(4): 409-434.

Scott, David and Weston, Ros (1998), *Evaluating Health Promotion*, Cheltenham: Stanley Thornes.

Shaw, Ian (1997), 'Evaluation in Health and Social Care', *Evaluation: The International Journal of Theory, Research and Practice* 3(4): 469-480.

Stinnett, A. *et al* (1996), 'The cost-effectiveness of dietary and pharmacological therapy for cholesterol reduction in adults', in Gold, M. *et al Cost-effectiveness in Health and Medicine*, New York: Oxford University Press.

Stockdale, Janet E, Gresham, Peter J, and Whitehead, Christine M.E. (1999), *Applying Economic Evaluation to Policing Activity*, Policy Research Series Paper 103, London: Home Office.

White, Michael and Lakey, Jane (1992), *The restart effect: evaluation of labour market programme for unemployed people*, London: Policy Studies Institute.