



Ministry of
JUSTICE

Towards a policy evaluation service

**Developing infrastructure to support the use of
experimental and quasi-experimental methods**

**Chris Creegan and Alan Hedges
National Centre for Social Research**

Ministry of Justice Research Series 8/07
August 2007

NOMS National Offender
Management Service
Working together to reduce re-offending

R D S
Research Development Statistics

Research, Development and Statistics exists to improve policy making, decision taking and practice in support of the Ministry of Justice purpose and aims, to provide the public and Parliament with information necessary for informed debate and to publish information for future use.

© *Crown copyright 2007*

Extracts from this document may be reproduced for non-commercial purposes on condition that the source is acknowledged.

First published 2007

ISBN 978 1 84099 088 1

Disclaimer

The views expressed are those of the authors and are not necessarily shared by the Ministry of Justice

Acknowledgements

A number of people at NatCen have played a part in this project: Meg Callanan, Andreas Cebulla, Norman Glass, Jon Hales, Riikka Hallikainen, Tim Knight, Clarissa Penfold and Susan Purdon. Thanks are due to them.

We would also like to thank the project board: Chloë Chitty (previously Home Office, now Ministry of Justice), Michael Bright (ESRC), Ricky Taylor (GSRU), Richard Dubourg (Home Office), Andrew Healey (was Home Office, now Ministry of Justice) and in particular Robin Elliott-Marshall (was Home Office, now Ministry of Justice) who managed the project on behalf of the board.

Sue Duncan (GSRU), Richard Bartholomew (DfES), Mike Daly (DWP) and Paul Wiles (Home Office) participated in an invaluable workshop to review the study findings and draft conclusions.

Members of SAGEM participated in discussion workshops that helped inform the project design.

Finally, we would like to thank the study respondents, both those who completed the web-based survey and in particular the 33 depth interview respondents. Their participation in the study has, for understandable reasons, been anonymised.

Contents

	Page
Summary	iii
1 Introduction	1
Research context	1
Aims and objectives	1
2 Methods	3
Sample design	3
Web-based consultation	3
Analysis	3
Other issues	4
3 Findings	5
The current situation	5
Developing the use of QEPO	7
Existing sources of support around QEPO methods	11
Need for further support around QEPO methods	12
How support should be provided	13
4 Implications of the study	17
Developing QEPO methods	17
Supply and demand	17
Defining support needs and criteria	19
Determining the location and structure	21
5 Conclusions	23

Summary

Introduction

This is a report of a scoping study concerning the development of infrastructure to support the use of experimental and quasi-experimental methods in quantitative policy outcome evaluation (QEPO methods). There has been increasing interest in developing ways of improving the efficacy of policy evaluation, and in particular debate about the use of experimental and quasi-experimental methods in the UK. A key aspect of this debate has been discussion about the possible value of establishing new infrastructure (one option being some sort of policy trials unit). The study was commissioned with cross-government support, led by the Home Office, Treasury, and the Economic and Social Research Council (ESRC). The study focused on five key objectives relating to infrastructure: need and demand; supply; delivery options; benefits and risks of options; and costs. The costs element does not form part of this report; a separate consultation process will take forward this and other issues associated with the creation and funding of a new policy evaluation service.

The core method of data collection was unstructured depth interviews, which were held with 33 key informants. This was supplemented by an open access web based survey to which 15 responses were received. The depth interview sample design was purposive. The sample was not and could not have been designed to be representative of the UK policy and research communities. Rather the objective was to include a range of key informants with known interest and expertise in the issues being considered. Respondents fell into three broad categories: users of research; providers of research and research experts. Respondents were identified in discussion with the project board, and snowballed from other respondents. The depth interviews were recorded and transcribed. They were analysed using Framework, a matrix based tool to enable systematic and rigorous analysis of qualitative data.

Findings

There was a broad spectrum of research method usage across government departments. At one end of the spectrum, in departments with little experience of evaluative research a limited range of methods had been used. At the other, in departments with long track records of evaluative research a wider range of research methods and more sophisticated approaches had been used. Departments with track records of using a wide range of non-QEPO methods also appeared to have most experience of using QEPO methods.

Respondents reported broadly similar models of decision-making in relation to policy testing. They had different experiences concerning the impetus to use QEPO methods, and cited a range of drivers including the Treasury and Treasury-based Government Social Research Unit (GSRU)¹, and the confidence of policy and research officials in the efficacy of QEPO methods. Experimental approaches had been encouraged by a combination of policy-makers, research commissioners and research contractors. User respondents felt there had been a push for more evidence-based policy-making since 1997, which had initially provided more scope to use experimental and quasi-experimental methods. Provider and expert respondents believed that government departments were generally more open to the use of QEPO methods than previously. An alternative standpoint was that the key issue was to promote good evidence-based policy-making, not QEPO methods specifically.

A number of issues emerged as central to the debate about QEPO methods including scope, time, ethics, expertise and knowledge and cost. Scope issues centred particularly on the efficacy of randomised controlled trials (RCTs). Time was raised as an issue in three respects: the amount of time needed to use QEPO methods; late consultation with researchers by policy-makers; and the adverse impact of quick

1 GSRU provides strategic leadership and support to the Government Social Research Service, including provision of specialist training and advice on experimental and quasi-experimental methods to analysts across government. See www.gsr.gov.uk for further information.

decision making. Two key ethical issues emerged: concerns on the part of delivery professionals and the sensitivities of randomising amongst particular populations. Alternative views were that ethical objections were sometimes used as an excuse or were a 'red herring'. Lack of knowledge and expertise on the part of politicians, policy-makers, research providers and research commissioners was perceived to be a key problem.

Existing support sources are available though not accessible or well known, most commonly GSRU, about which there were mixed views. The prevailing view was that there is a need for infrastructure to support the use of QEPO methods, though both ambivalence and opposition to this view were also evident. Organisational location was seen as a critical issue, in particular whether or not a service should be located within government, but there was some support for a unit at arms length of government. A virtual network was thought to be insufficient to meet the needs identified. Various potential support roles were identified, including training, advice and support and championing QEPO methods. The importance of QEPO expertise and policy awareness was emphasised. Two roles, those of regulator and provider (in the sense of conducting evaluations), were ruled out.

Implications of the study

There is support for encouraging and enabling wider use of QEPO, though the efficacy and ethical basis of QEPO techniques remain contested. This suggests a need to focus on optimising rather than maximising QEPO use, to ensure the appropriate use of QEPO methods, and for a service to include a 'triage' role to assess the viability of approaches at the outset of projects. If QEPO is to develop along these lines a proactive approach is needed. This study provides evidence of interest and in some cases considerable enthusiasm, but little evidence of momentum or innovation.

There is support for developing infrastructure to provide the impetus that is needed and provide support in relation to the use of QEPO methods. However, there is also a need to acknowledge and where possible respond to arguments from those who are not persuaded by the case for doing so when moving towards the creation of any new policy evaluation service. This underpins the formulation of the study's recommendations about the approach and design of a new service.

Since the current situation is static and there are multiple perceived barriers and disincentives, a proactive approach is needed on the demand and supply sides. On the supply side it is clear that research commissioners are key and need to be better informed and more confident about QEPO methods. Providers of research services need to be encouraged to tool up for increased demand. A shared understanding is needed between policy-makers, commissioners, providers and practitioners about three central issues: the place of QEPO within policy evaluation; the roles of prospective and retrospective QEPO; and the functions of different types of QEPO technique.

The overriding need is for a policy evaluation service that facilitates QEPO development through information, advice and training. A key challenge will be to balance breadth in relation to all policy evaluation methods with depth in relation to QEPO. This means addressing the reported skills deficit in QEPO methods within a problem/question driven context rather than a technique-driven context, which implies, for example, that RCTs are a 'gold standard'. A service that is able and willing to say that what is needed is a non-QEPO approach is likely to have greater and wider credibility.

A multiplicity of functions will be necessary to deliver the support needed to create impetus, which fall into four categories: promotion and training; knowledge and information; advice and support; and development and innovation (described further in Figure 1).

Figure 1 Overview of possible functions of a policy evaluation service

Promotion and training	<ul style="list-style-type: none"> ● championing and promoting appropriate use of QEPO ● facilitating training and capacity building
Knowledge and information	<ul style="list-style-type: none"> ● facilitating the build-up of expertise and resources ● acting as an information clearing house
Advice and support	<ul style="list-style-type: none"> ● delivering an advice function on QEPO use ● providing practical guidance/ help with implementation
Development and innovation	<ul style="list-style-type: none"> ● facilitating use of secondary data and meta analysis ● enhancing standards and providing quality assurance ● clarifying and codifying QEPO approaches

There were conflicting opinions about the benefits and risks of various location options. On balance, despite the fact that the dominant view was for a structure within government (probably alongside GSRU), the authors' view is that the best option would be an external location, either an academic or an independent research provider, combined with a virtual network. The organisation will need to have a broad range of existing methodological and substantive expertise and demonstrate commitment to the development of QEPO methods, both intellectually and financially. It should draw on expertise from within government and have clear links with the ESRC's National Centre for Research Methods (NCRM) and its node network should be established.

Conclusions

The study indicates that there is support for widening the use of QEPO methods. Though strong differences do exist in the wider research community, these should not preclude proper consideration of the use of QEPO methods in policy evaluation. There is scope to build greater consensus about the role of experimentation within policy evaluation. Key to this will be collaboration within and outside government involving ministers, senior policy officials, the Treasury, research commissioners and providers of research services. The focus of collaboration should be the creation of an agenda, which is premised on the optimum use of QEPO generally and experimentation in particular.

The barriers identified underline the need for intellectual, practical and financial investment in the development of QEPO methods. The current apparent lack of momentum provides a strong case for the development of infrastructure, based on shared dialogue between key stakeholders. The success of infrastructure will, of course, depend on adequate financial investment. However, the research suggests a range of other factors will be significant. First, a proactive and entrepreneurial approach will be needed to drive the QEPO agenda forward and establish the service's credibility across institutional and methodological boundaries. Second, the provision of a confidential service will be vital to address concerns about sharing sensitive information outside government. Third, advice and support should initially be free at the point of delivery so as not to deter those sceptical about the use of QEPO methods.

Recommendations

Developing QEPO methods

- Steps should be taken to optimise the use of experimental and quasi-experimental methods in policy evaluation.
- Optimisation of the use of QEPO should be sensitive to policy context and involve a combination of methods where appropriate.
- Project needs should be analysed at the outset in order to review the proposed evaluation methods and the role of QEPO.

Supply and demand

- Steps should be taken to foster dialogue and understanding about the use of QEPO between policy-makers, research commissioners, research providers and practitioners.
- Dialogue should focus on the place of QEPO within policy evaluation, the roles of prospective and retrospective QEPO and the functions of different types of QEPO technique.
- The demand for QEPO should be stimulated by providing clear advice for ministers, senior policy officials and the Treasury about the role of experimental and quasi-experimental methods within the spectrum of available evaluation options.
- The supply of QEPO should be facilitated by ensuring that research commissioners are properly informed through the provision of support, advice and training, and that research providers are encouraged to invest in the development of skills, where appropriate through consortia arrangements.

Support needs and criteria

- Support for the development of QEPO methods needs to be provided through the provision of information, advice and training in a broad evaluative context.
- The functions of a policy evaluation service should focus on four key areas: promotion and training; knowledge and information; advice and support, and development and innovation.
- The credibility of any policy evaluation service will need to be built carefully through the dissemination of practical messages and by addressing perceived ethical concerns about QEPO methods.
- The provision of initial advice on QEPO methods by a policy evaluation service should be free, but a charging system for ongoing project support should be introduced.

Location and structure

A policy evaluation service should:

- be independent and located outside government in order to maximise involvement, with appropriate arrangements for confidentiality;
- have strong links to the ESRC's National Centre for Research Methods and its node network;
- be provided from within an academic or independent research organisation with a clear commitment to developing the use of QEPO methods and a broad range of methodological and substantive expertise;
- be underpinned by strong leadership, entrepreneurial skills and the secondment of personnel to maximise opportunities for exchanging ideas and expertise;
- include both physical and virtual infrastructure.

1 Introduction

This is a report of a scoping study concerning the development of infrastructure to support the use of experimental and quasi-experimental methods in policy evaluation. The study was commissioned with cross-government support, led by the Home Office (HO), the Treasury-based Government Social Research Unit (GSRU), and the Economic and Social Research Council (ESRC). It was commissioned following discussion with Government Social Research (GSR) representatives from across departments, and with the support of the Strategic Advisory Group on Experimental Methods (SAGEM)², a cross-government group led by the Home Office, which involves experts from inside and outside government.

Research context

The study took place in the context of increasing interest in developing ways of improving the efficacy of policy evaluation and debate about the use of experimental and quasi-experimental methods in the UK (see, for example, Greenberg and Poole, 2007). A key aspect of this debate has been discussion about the possible development of infrastructure to raise standards and support researchers engaged in policy evaluation within government, universities and other research bodies. One option that has been discussed is the establishment of a central unit to support the evaluation of policy trials (see, for example, Bird, 2004).

Such discussion has been based on the premise that randomisation has a crucial contribution to make to robust and cost-effective policy evaluation. However, the focus of this study was on the broader range of experimental and quasi-experimental methods in the quantitative evaluation of policy outcomes, which the authors have termed QEPO³. It is important to acknowledge at the outset that experimental methods, and QEPO more generally, are only part of the policy evaluation jigsaw. For example, policy evaluation may involve a process element or qualitative evaluation. Indeed, these dimensions inevitably arose during the study, both in the authors' own thinking and that of respondents. However, they were not the direct focus of this study.

Aims and objectives

The central aim of this study was to inform the possible establishment of a new policy evaluation service. It therefore focused on five key objectives relating to the development of service infrastructure:

- Need and demand: what are the perceived needs of the UK policy and research communities to support the use of QEPO and what demand is there for support to be provided?
- Supply: what are the possible functions of infrastructure to support and develop the use of QEPO methods?
- Delivery options: what types of structure could be used to achieve these ends? Could existing structures or models of support be used?
- Benefits and risks: what are the perceived benefits and risks associated with the various options?
- Costs: what are the likely costs of the main delivery options?

This report deals with the first four of those objectives. The costs aspect was covered in a separate report to the project board. A separate consultation process will take forward the detailed implementation issues associated with the creation and funding of a new policy evaluation service.

2 The aim of SAGEM is to act as a discussion forum, to advise on the strategic development of experimental and quasi-experimental methods, to share expertise and knowledge, and help raise research standards.

3 For a more detailed description of these methods see Chapter 7 of the Magenta book www.policyhub.gov.uk/magenta_book/

There is also a growing literature on experimentation and randomisation in social sciences and public policy (see, for example, Jowell, 2003). However, it was not the purpose of this study to review or summarise the arguments in that literature. Rather, its purpose was to explore the issues related to the above objectives with a range of key informants in order to move the debate about infrastructure forward and help inform future decision-making within government and research funding bodies.

2 Methods

The study was qualitative in design. The core method of data collection was unstructured depth interviews, which were held with 33 key informants (Topic guides attached as Appendix 1). This was supplemented by an open access web-based survey with a series of open questions (Questions attached as Appendix 2).

Sample design

The depth interview sample design was purposive and respondents fell into three broad categories: users of research; providers of research; and research experts who did not fall neatly into one category. A purposive sample is not designed to be statistically representative, but rather to capture a diverse range of characteristics relevant to the study. The sample included 19 user respondents from a range of government departments. They included research officials, policy officials and others with an interest in the use of experimental and quasi-experimental design. Fourteen provider and expert respondents were largely, though not exclusively, drawn from outside government. They included academics and research providers with interest and expertise in these methods mainly in the UK, but also from the US. Respondents were identified in discussion with the project board, and where appropriate snowballed from other respondents.

It is vital to acknowledge that this sample was not, and indeed could not have been, designed to be representative of the UK policy and research communities. First, it was limited to those who were known to have an existing interest in QEPO either as potential users or providers of policy evaluation, and those who were known experts. Second, it was obviously not possible to include all known providers and experts, though purposive selection involved seeking to include informants from a range of substantive disciplines including, for example, criminology, education and health. Third, it was not possible to include policy and research officials from all government departments. The sampling strategy was, therefore, to include officials from both large and small departments, the Scottish Executive, the Welsh Assembly Government (WAG), and departments with different levels of expertise and experience in relation to QEPO. Participating departments included the Home Office, the Department for Work and Pensions (DWP), the Department for Education and Skills (DfES), the Department for Constitutional Affairs (DCA) and the Health and Safety Executive (HSE). Finally not all of those who were approached both inside and outside government agreed to take part.

Web-based consultation

In order to ensure that anyone not included in the sample with a strong wish to participate in the study could do so, a web-based consultation exercise was used which included a survey of open questions relating to the core study objectives. This was open access and publicised widely via the National Centre for Social Research (NatCen), GSRU, ESRC, the Social Research Association (SRA) and the UK Evaluation Society (UKES) websites and via the Royal Statistical Society. A total of 15 responses were received from a mixture of government officials, academics and others. These responses were summarised and considered alongside the depth interview data. This is clearly a small number of responses. Although the survey was publicised widely, the timescale for responses was tight and there was no opportunity for follow up.

Analysis

The depth interviews were recorded and transcribed. They were analysed using Framework: a matrix based tool developed by NatCen to enable the systematic and rigorous analysis of qualitative data. Three thematic charts were created to map respondents' views and experiences in relation to the current use of QEPO methods, widening the use of QEPO methods and the development of infrastructure to support the use of QEPO methods. The interview and survey data were summarised and synthesised in the charts to inform the analysis of findings, which is contained in section 3 of this report.

Other issues

There were three other important aspects to the study design. First, group discussions were held with SAGEM members at the beginning of the study to inform topic guide design for the depth interviews. Second, following completion of the interviews a group discussion was held with the project board members and senior government research officials to review headline findings and consider draft conclusions. Third, the authors approached the study as consultants as well as researchers. In that sense they brought their own perspectives and experiences which were drawn upon when weighing up the evidence and exploring possible solutions. The conclusions are, therefore, not only grounded in the data. They are also born out of a creative fusion between commissioners, participants and researchers.

3 Findings

The current situation

Evaluation approaches

Unsurprisingly, the study indicated that there was a broad spectrum of research method usage across government departments. In part this appeared to reflect the extent to which a culture of evidence-based policy was embedded. At one end of the spectrum, respondents from departments with little experience of evaluative research reported that a limited range of methods had been used, for example small-scale *post hoc* surveys and qualitative studies. Some doubts were expressed about the rigour and robustness of previous evaluations.

At the other end of the spectrum, respondents from departments with long track records of evaluative research typically reported the use of a wider range of research methods and more sophisticated approaches. These included large-scale pilot/control group surveys, combining analysis of survey and administrative data, econometrics, economic modelling and extensive use of large-scale qualitative research. Whilst some concerns were expressed about rigour and robustness, in general respondents felt that the research that was undertaken was of good quality.

There did appear to be a conscious push to innovate and drive up standards across departments, both in terms of raising the quality, and extending the breadth of research. In the Home Office for example a project quality review board has been established. One influential factor was said to be the arrival of key individuals with strong views about evidence-based policy.

Use of QEPO methods

The study suggested that the extent to which departments had experience of using QEPO methods could broadly be mapped on to the above spectrum. Departments with track records of using a wide range of non-QEPO methods also appeared to have most experience of using QEPO methods. Amongst these departments the Home Office, the Department for Work and Pensions and the Department for Education and Skills had used quasi-experimental methods quite widely, including propensity score matching. Both the Home Office and DWP have used randomised controlled trials. Within the Home Office, RCTs have been used to evaluate a range of interventions, including the impact of providing information on public attitudes to sentencing (see Chapman *et al.*, 2002), Restorative Justice (RJ) (see Shapland *et al.*, 2004, 2006 and 2007), and the short-term effectiveness of the Enhanced Thinking Skills (ETS) programme for convicted offenders (generally known as the Treatment Change project⁴). Within DWP the two large-scale examples are the Job Retention and Rehabilitation Pilot (JRRP) project (see Natarajan *et al.*, 2005) and the Employment Retention and Advancement (ERA) demonstration project (the implementation of which had involved officials in other departments) (see Hall *et al.*, 2005). An RCT design had been used more recently on a smaller-scale project concerning pensions advice in the workplace (Leston and Watmough, 2005).

Respondents from the Department for Constitutional Affairs and the Health and Safety Executive reported limited use of QEPO methods. The DCA had conducted a small-scale RCT on debt advice, while the HSE was conducting an evaluation of a new initiative using propensity score matching for the first time. In Scotland only quasi-experimental approaches have been used to date in relation to evaluations of fast tracking hearings, secure care and an initiative involving intensive interventions with young people with

4 A large Home Office-funded project entitled 'Evaluation of HM Prison Service accredited cognitive skills programmes: Treatment Change Project'.

disruptive behaviours. The WAG, however, though still in the relatively early stages of building up research and evaluation expertise, has supported a programme evaluation using an RCT (Stop Smoking in Schools Trial (ASSIST), a peer leadership-type programme). This involved a cluster randomisation of secondary schools across Wales and the south west of England funded by the WAG and the Medical Research Council (see Starkey *et al.*, 2005). Two further evaluations using RCTs are under way: firstly, free breakfast provision in primary schools (see Tapper *et al.*, 2007) which involves a cluster-randomised trial in primary school (the school is the primary unit) and secondly an exercise referral programme⁵ which involves individual randomisation (the National Institute for Health and Clinical Excellence (NIHCE) recommended that any new programme should only be introduced in the context of a trial).

This pattern could suggest two different scenarios. On the one hand it could be inferred that the use of QEPO methods is part of an evolutionary process, i.e. it represents a stage of development that is reached when a culture of evidence-based policy-making has been firmly established and sophisticated methodological expertise has been developed. On the other it could suggest that developing the widest possible range of evaluation methods including QEPO encourages such a culture. Either way the evidence underlines that fostering a culture of evidence-based or informed policy-making and widening the scope of evaluation methods are interdependent developments.

Decision-making on policy testing methods

Respondents reported broadly similar models of decision-making in relation to policy testing. In the first instance, policy officials met with government researchers to agree the most appropriate methodology with ministers giving final approval. The project specification would then go to potential tenderers, either giving a broad indication that a particular approach was expected, or in some cases explicitly asking for QEPO methods (including as one element of a mixed method approach). Policy officials sometimes commissioned research themselves, with researchers only involved in an advisory capacity, though it was not suggested that this approach was less likely to lead to QEPO methods being used. For example, the HSE has several quasi-experimental evaluations, some of which are policy-led.

Respondents had different experiences concerning the impetus to use QEPO methods, and suggested a range of drivers including the Treasury and GSRU, and the confidence of policy and research officials in the efficacy of experimentation and randomisation. However, these drivers were not one-dimensional or mutually exclusive. For example, the Treasury was said to have both encouraged the use of experimental methods through a general push for robust evaluation and explicitly demanded a randomised approach in one particular instance.

It was reported that experimental approaches had been encouraged and suggested by a combination of policy-makers, research commissioners and research contractors. For example, in one department an RCT had been proposed by researchers in response to a policy-led desire for high quality evaluation because a particular intervention was high profile (and presumably expected to work). In the same department research contractors had suggested a randomised design after commissioners had given some indication that it was a desirable approach. In cases where QEPO methods (specifically RCTs) were seriously considered it appeared that the decision-making process could become quite protracted. Respondents referred to a long process of internal meetings and consultation with experts. A range of factors would be weighed up including the potential for negative publicity, which was said to have been a key factor in one decision not to go ahead with an RCT.

⁵ The Evaluation of National Exercise Referral Scheme (National Assembly for Wales) is being led by Dr Simon Murphy and Professor Laurence Moore at Cardiff University and is due to report in August 2009.

Recent use of QEPO methods

In general, user respondents recognised that there had been a push for more evidence-based policy-making since 1997⁶, and that initiatives had been implemented to facilitate better policy-making during this period (see Campbell *et al.*, 2007). It was felt that this push had initially provided more scope to use experimental and quasi-experimental methods. The cyclical nature of policy-making meant that more recently there had been fewer new policy initiatives, and nervousness about commissioning evaluation had crept in.

Respondents with experience of RCTs suggested that initially a lot of mistakes had been made. It was felt that departments had previously not been good at shared learning, though this was now emerging. Such learning, combined with the use of US expertise, meant that the capacity to carry out QEPO research was increasing, though it still appeared patchy.

Amongst those who expressed a view, the main impression was that the findings of recent RCTs in the UK had demonstrated limited differences in outcomes for those on and off programmes, resulting in scepticism amongst policy-makers (an experience which was familiar to US colleagues). It was also suggested that the failure of RCTs could be attributed to poor implementation of randomisation, relating for example, to sample contamination and insufficient sample sizes. Such problems were in part attributed to lack of experience, albeit lessons had been learned in some departments.

Amongst provider and expert respondents, it was believed that government departments were more open to the use of QEPO methods than previously, for example, in relation to the use of quasi-experimental methods to exploit longitudinal data sources. The general perception, though, was that QEPO methods were still under utilised. An alternative less common standpoint was that the key issue was to promote good evidence-based policy-making generally through the use of mixed-methods and systematic reviews, not QEPO methods specifically.

In some cases the view that QEPO methods were underutilised was combined with a desire for such methods to be used more widely in policy evaluation, though it was acknowledged that they would never be appropriate for addressing all research questions. A further key issue that emerged in relation to the utilisation of QEPO methods was the need to consider their use both prospectively in relation to new interventions, and retrospectively through the use of secondary data and meta-analysis re-analysing data from past pilots and trials. This would have two key advantages. First, it would be extremely cost-effective. Second, it would enable methodological innovation and provide opportunities to develop and exploit new techniques.

Developing the use of QEPO

Various issues emerged as central to the debate about the development of QEPO methods. These include scope, time, ethics, expertise and knowledge and cost.⁷ The evidence in relation to each of these issues, will now be considered.

Scope

A range of views was expressed about the merits and advantages of QEPO methods, particularly RCTs. One view was that RCTs represented the 'gold standard' or the 'Rolls Royce' of evaluation methods, which should always be used if feasible. In some instances RCTs were described as providing a robust testing mechanism without clear explanation of why this was thought to be the case. The use of the 'gold standard'

6 It should be acknowledged that neither experimental design nor evidence-based policy-making are new phenomena in government. However, respondents in this study focused on recent developments, i.e. post 1997.

7 For a more detailed discussion of the issues associated with randomisation and quasi-experimental methods, see Chapter 7 of the Magenta book www.policyhub.gov.uk/magenta_book/

label was generally justified by reference to RCTs' ability to provide an unbiased estimate of effect, with all other potential influences filtered out. It was argued that this provided greater certainty and confidence in evaluation findings, thus making the results easier to defend. However, it was not necessarily suggested that they should or could be used in isolation even by those who most strongly advocated their use. There was also some antipathy to the use of the term 'gold standard', because it implied rigidity, technical purism and an 'evangelistic' standpoint.

The 'gold standard' label was challenged in two key respects. First the unbiased estimate of effect assertion was contested, including by advocates of RCTs. It was argued that it is still necessary to make certain assumptions when using an RCT, just as it is in relation to the use of quasi-experimental methods. It was also suggested that bias could be introduced into an RCT if randomisation and the implementation of the policy were not being applied consistently on the ground. Essentially it was suggested that RCTs could be as prone to bias as other methods if they are not carried out properly and that they are most likely to go wrong during the implementation phase.

Second it was argued that RCTs could provide a measure of impact, without necessarily shedding any light on how or why a policy intervention had worked or not. This position was most strongly advanced by those who explicitly made the case for a mixed method approach to evaluation, though as has already been acknowledged, advocates of RCTs did recognise the need for methodological pluralism. It was recognised that the same argument could be levelled at the use of other quantitative methods when attempting to measure impact.

Quasi-experimental methods, on the other hand, were sometimes described as 'silver standard', a step up from a standard approach, but not as robust as RCTs. The underlying reason for this view was that such methods are based on assumptions about causal factors and the criteria on which people are matched. However, they were thought to be particularly useful in some situations (e.g. in relation to community interventions), though care was needed to interpret results in relation to the delivery of the mechanism and the local context.

Time

Time was raised as an issue in three key respects. First, the amount of time needed to use QEPO methods, especially RCTs, was perceived as a potential disadvantage. It was generally believed that considerable time was needed to run an RCT properly, more than with the use of other methods. This was seen as a problem in the context of the desire of policy-makers for quick findings and the 'short time horizons' of ministers, which were widely cited as barriers to commissioning RCTs. On the one hand this was seen as inevitable and there was some sympathy for the need for timely information, including interim reports and early signs of progress. On the other, there was also a degree of cynicism about politicians simply wanting quick results for political gain.

Ultimately, this was not seen as something that could be addressed without a radical sea change in the process of government decision-making. For example, one respondent with experience of government research talked of the need to move away from the current 'fire fighting model' of policy-making, towards a more strategic, longer-sighted approach, for example by having a five-year research plan at the start of parliament. However, there were also pragmatic suggestions about how the situation could be improved without such radical change. Typically these focused on ensuring that policy-makers consulted government researchers earlier in the policy-making cycle so that there were greater opportunities to consider and utilise the widest possible range of evaluation methods including experimentation. This underlined the need to increase policy-makers' and ministers' understanding and appreciation of research and make a business case for integrating research into policy development.

Second, opportunities to use QEPO methods were missed because policy-makers did not consult research colleagues about the evaluation of a new policy until it was already up and running or later than was ideally desirable, though in some departments this was said to have changed. This problem was linked to a common perception that some policy-makers and ministers lacked a proper understanding of research methods and the need for good evidence to support policy-making. The result was that interventions were sometimes funded without proper thought being given to evaluation.

Third, it was argued that the practice of policy-making often militated against QEPO approaches because of the pace with which decisions were made. The result was pressure to deliver results quite early in a programme evaluation such as Sure Start, even though it was necessary to allow time for the programme to operate as intended before examining causal impacts and assessing cost-effectiveness.

Ethics

Ethical concerns were raised in connection with QEPO methods. Though concerns were raised in relation to quasi-experimental methods, for example in relation to the linkage of sensitive data, they were primarily about RCTs. Two key issues were identified. First, it was argued that delivery professionals and practitioners asked to implement RCTs would be reluctant or may even refuse to do so on ethical grounds. The clearest example of this was in relation to judges and sentencing in both England and Scotland, because it was believed that they would refuse to implement randomisation.⁸ Another example was probation staff who may be concerned about non-intervention in a situation, which raised issues of public protection.

Provider and expert respondents emphasised the importance of talking directly to those responsible for randomisation of delivery to address any ethical concerns. Two successful examples of this were referred to. First, the ERA evaluation involved the creation of 'technical advisors', who worked with jobcentre staff to ensure that randomisation was implemented properly. Second, a key learning point from the cluster-randomised trial of the primary school free breakfast initiative in Wales, was that "winning the hearts and minds of policy officials and key stakeholders had been critical to the successful implementation of the trial" (Hale *et al.*, 2006).

The second ethical concern raised was that ministers would have reservations about commissioning an RCT because of the sensitivities associated with randomising delivery and might even refuse to allow an RCT to be commissioned. This problem was more pervasive and cited as a potential barrier by respondents across departments, which could inhibit government researchers and policy-makers from putting forward a randomised design. It was also argued that it was necessary to think very carefully about the population group when considering an RCT. New Deal for Disabled People was given as a case in point where it was felt on reflection that randomisation was always likely to be sensitive and emotive with such a vulnerable population combined with a vocal campaign lobby. Within the DfES there was nervousness amongst policy-makers about randomising delivery to schoolchildren.

An alternative view was that ethical objections were sometimes used as an excuse and that there was greater scope for the use of randomisation. For example, it was argued that the public had more intelligence than they were given credit for and that in a climate of uncertainty about the best model of provision, a trial represented the most ethical option, and that in this sense experimentation offered fairness. A more exceptional view still was that ethical objections were a 'red herring'.

⁸ Of course, this need not always be the case. In the Restorative Justice evaluation there was senior judicial support, for various reasons, including recognition that RJ was not something participants were entitled to, and that the aim of developing knowledge on effectiveness justified withholding it from some.

Knowledge and expertise

Lack of knowledge and expertise on the part of politicians, policy-makers, research providers and research commissioners was perceived to be a key problem, particularly in relation to experimental methods. This could result in negativity and even hostility.

Research commissioners and providers

The general impression was of isolated and dispersed expertise in the UK research community. As one respondent put it, there are "pools of knowledge in a desert of ignorance". Lack of expertise amongst government research commissioners was a commonly cited barrier to the use of QEPO, though amongst user respondents this view was held more strongly in some departments than others. Whilst some were reasonably confident that there was sufficient knowledge about quasi-experimental methods, there was a perceived skills gap in relation to the use of experimental methods. Others believed that the skills gap applied to both. A consequence of this was that government researchers did not suggest or promote the use of QEPO methods, either because of insufficient awareness or lack of confidence.

The main explanation offered for the shortage of experimental skills, particularly by academic respondents, was the content of research degrees and masters courses in the UK. Curricula were perceived as weak in relation to quantitative methodologies generally and experimental methods specifically, as a result of which they were ignored by PhD students. And concern was expressed about the prevalence of anti-positivism in the UK research community. Consequently, it was suggested that graduates had neither the knowledge nor appetite for QEPO. It was suggested that the more widespread use of QEPO methods could be facilitated by increased coverage in syllabuses of research degrees and placements for PhD students in government departments. There was some evidence that steps had been taken to enhance coverage of this at a limited number of universities such as Oxford.

However, it was recognised that this was a long-term solution and there were also calls for more professional training on QEPO methods to be available to government researchers. Bespoke training was preferred because the timing of academic courses was not seen as sufficiently flexible. Workshops where QEPO experts could disseminate skills and experience to government researchers were also suggested, as were secondments of expert staff to less experienced departments.

Another potential source of learning was thought to be experts in the US. The DWP RCTs were cited as an example of the value of UK researchers learning from their US counterparts, and further joint working of this nature was thought to be worth promoting. Moreover, American respondents (perhaps predictably and with understandable self-interest) identified this as providing further potential benefits. There was even a suggestion that US firms should be encouraged to set up a branch in the UK, but the general preference seemed to be developing practical skills within the UK research community. It was emphasised, however, that such measures would, in practice, need to be combined with stimulating demand and creating a market for QEPO research.

Some references were made to the relationship between lack of expertise amongst both commissioners and providers and lack of demand. One suggestion was that UK researchers lacked skills simply because hitherto they had not been expected to have them. This view was also implicit in other messages about the relationship between lack of practical expertise and lack of real opportunities to use randomisation.

Policy-makers and politicians

The knowledge and expertise of policy-makers and ministers was criticised for a variety of reasons. These included not having a real appreciation of how research should be used to support policy-making, lack of knowledge to properly weigh up the pros and cons of different approaches, and risk aversion and

nervousness in relation to unfamiliar methodology. Non-QEPO methods, for example qualitative work focusing on process and *post hoc* surveys, were seen as safer options. The result was a lack of demand for the types of high quality evidence that some QEPO methods were seen to offer. In some cases, policy-makers were said not to be consulting government researchers at all about research needs.

Different views were expressed about the basis of this lack of knowledge. For example, one suggestion was that decisions were made for essentially political reasons, whilst another was that policy-makers often lacked a basic understanding of research methods or were statistically illiterate. The latter view was more common and recent research was said to have revealed poor understanding of research methods amongst policy-makers, particularly QEPO methods. Nevertheless, it was also reported by respondents with experience of running RCTs that policy-makers had been very supportive, particularly once issues had been explained to them.

There were a range of suggestions about ways of increasing the knowledge and changing the attitudes of policy-makers and ministers. The first was to develop practical advice in the form of an equivalent to the Magenta book for policy-makers⁹. This was envisaged as a manual, which explained the pros and cons of different methods and their appropriateness in different contexts. One respondent referred to a process in the US, called 'blue print', which involves star ratings of different methodologies. The second was a suggestion that building knowledge of research methods into job descriptions and professional competencies for policy-makers, could act as a spur to them seeking out more information and training about QEPO methods.¹⁰

Another view was that policy-makers and ministers would only start to take QEPO methods seriously as an option if there were practical examples of RCTs¹¹ that have been successfully run in the UK. It was argued that this could partly be achieved by doing a better job of publicising the RCTs that had already been conducted in the UK. There was a view that government researchers simply needed to get better at explaining the benefits of QEPO methods to policy-makers and ministers and be more confident about putting them forward as options. In turn it was argued that policy-makers didn't always have the confidence to persuade ministers of the need for research or the case for taking a particular approach. In contrast it was suggested that by pumping some initial funding into the conduct of a few relatively small-scale RCTs, additional evidence of implementation in practice would be generated.

Cost

A range of views was raised about cost-related issues. It was felt that RCTs, in particular, were expensive. For example, in one instance an RCT had been turned down because it had been possible to fund two non-randomised evaluations for the same price from the same provider. However, it was also argued that there were misunderstandings about costs. For example, randomising delivery was not the most expensive aspect of the ERA evaluation and QEPO methods were not necessarily more expensive than other methods, and only really high compared with analysing existing data. Whilst it was argued that cost should not be seen as a barrier given the relatively generous funding of policy evaluation, costs were seen as a barrier by respondents in smaller departments and in the context of more stringent spending.

Existing sources of support around QEPO methods

The overall impression of respondents in this study was that existing sources of support are available though not accessible or well known. In general, users felt that there was someone they could go to for support, but beyond this there were differences in terms of their knowledge of specific sources of advice. At one

9 www.policyhub.gov.uk/magenta_book/

10 This suggests limited awareness of the developments which have taken place in relation to Professional Skills for Government (www.psg.civilservice.gov.uk). The Professional Skills for Government (PSG) agenda has 'analysis and use of evidence' as one of its four core skills for all civil servants (there are two additional skills for senior civil servants).

11 This argument could be applied to other QEPO methods.

end of the spectrum there were users with little or no experience of commissioning QEPO work and who did not have close links with bodies of expertise such as GSRU or SAGEM. Typically these respondents commented that if they needed to know more about QEPO methods they thought they could ask for help or signposting to help from colleagues or researchers in other departments. Others cited a particular individual or organisation to whom they would turn. At the other end of the spectrum those with experience of running RCTs were familiar with a range of support sources. The various sources of support referred to included GSRU, individual government researchers, UK research contractors and UK academics including those involved in medical research, and US contractors.

The most commonly cited source of possible support was GSRU, though comments were not always entirely positive. One view was that the use of RCTs had been pushed quite hard, possibly at the exclusion of giving help and support around other methods. Another was that the support that had been provided by GSRU had been largely generic, and that support tailored to research questions (possibly from academics) might have been more helpful. The changes over time in the structure and function of support for government social researchers, in particular the demise of the policy evaluation division which existed as part of the Centre for Management and Policy Studies until 2002, appeared to have changed people's perceptions of GSRU as a source of support. There seemed to be some uncertainty about where things stood now and no recent examples of support were discussed. GSRU respondents acknowledged that they currently had limited resources to deliver this kind of support, and acknowledged it was largely reactive. It was also felt that those who approached GSRU for support tended to be 'the converted', rather than those with little or no current expertise in QEPO methods.

Need for further support around QEPO methods

The prevailing view amongst all respondents was that there is a need for infrastructure to support the use of QEPO methods. Two interconnected perspectives associated with this view were a desire for QEPO methods to be used more widely and the belief that perceived barriers to their use needed addressing. The key underlying factors for this belief were the need to raise quality, provide technical support and develop training. A more cautious approach was that further support should be provided but that the investment of resources in infrastructure should be limited, at least initially when demand was likely to be low. Concern was also expressed that infrastructure might have limited impact either because of insufficient investment or because infrastructure focused solely on QEPO methods would be isolated or marginalised.

However two more exceptional views were those of ambivalence and opposition to the development of infrastructure. Ambivalence was characterised partly by a view that additional support was not needed and partly by a view that if there were a demand for support it would be generated by the market (though it was accepted that infrastructure would be needed for things to happen more quickly). The former view was mainly expressed by those in departments with existing expertise and who already had very good links to support. Related to this there was also a view that while there might be a need for support infrastructure, its only real role would be to bring together existing sources of support. Nevertheless, it was accepted that there was some advantage to creating a body which provided a single source of support.

Opposition was based on a range of arguments. First, it was argued that there were more pressing issues to be addressed such as the promotion of good quality research and evidence-based policy-making generally. One dimension of this argument was that the investment required could not be justified when research and evaluation budgets were being cut. Second, it was suggested that infrastructure would push QEPO methods too hard at the expense of other approaches. A key aspect of this argument was that there is currently a lack of methodological cross fertilisation in policy evaluation. Thus it was suggested that what is needed is an initiative which binds together different methodological approaches rather than placing a particular emphasis on QEPO. Third, it was argued that there was insufficient demand to justify the development of infrastructure, either for the use of QEPO methods or for advice and support about their use. This again made the

investment of resources in such a venture questionable in the absence of demand being generated. It was also suggested that existing uneven demand would make it difficult to secure sufficient sign-up.

How support should be provided

Location of support infrastructure

Organisational location was seen as a critical issue for the potential success or failure of infrastructure. An exceptional view, held mainly by users, was that the location of infrastructure in organisational terms was irrelevant. What mattered was that it was staffed by people who knew what they were doing, and who could provide the support needed. For most respondents, however, the key dimension was whether or not it should be located within government. The pros and cons of different options are summarised in Table 1.

Within government

The main argument advanced in favour of locating infrastructure within government was that advice and support needed to be delivered by people who understood the political environment, and the decision-making context within government. This was underpinned by a perspective that it would need to be credible amongst policy-makers and ministers, and that its influence and authority would be enhanced by proximity to those with spending capability.

Amongst those who believed that it should be located within government, there were different views about the best precise location. There was some sympathy for having it within one department, largely because there would be an existing pool of expertise to build on, combined within an understanding of policy-making. However, in general, this option was ruled out on the basis that researchers from other departments would be very reluctant to go there for advice, particularly if they considered themselves to be knowledgeable in QEPO methods. The perceived tribalism and insularity of departments were both factors here; as one respondent put it things are still very 'departmental'. Questions were also raised about whether staff in the chosen department would be able to devote time to helping those in other departments, alongside their 'day job'.

More favoured options were placing it either within or alongside GSRU. Association with the Treasury was identified as an important source of credibility and demand push. There were contrasting views about the option of placing it within GSRU. On the plus side it already had an established, cross-government remit, and was recognised as having a degree of autonomy. However doubts were raised about its impact and credibility and whether it was sufficiently connected to policy-making. It was also recognised that it did not currently have the expertise, capacity or resources to fulfil the role.

Outside government

Those favouring an external location argued that if it were independent of government, it would have greater intellectual freedom and could make recommendations, without being subject to political pressures or the 'whims of the ministers'. It was also felt that infrastructure located outside government would be seen to have greater external credibility, that independence would provide greater stability in the longer term and that it would be more accessible to potential users. These views were particularly, though not exclusively, advanced by academic respondents. Such counter-arguments about independence created a dilemma for some respondents: if infrastructure was at arms length it would enable those within it to comment freely about the effectiveness of programmes, but on the other hand the distance from government would dilute its ability to impact directly on policy-making.

A range of concerns was raised about locating infrastructure within academic institutions, particularly by government researchers. Academics were often seen to be too theoretical, naïve about the political environment, and more interested in getting papers published than providing useful, practical advice.

Academic institutions were also characterised as poorly managed, weak at communicating with the outside world and having limited influence over key decision-makers. On the other hand it was recognised that the key benefit of locating it in an academic institution was that it would be able to build on existing structures and expertise, though this was not necessarily considered enough to offset other concerns.

Similarly locating infrastructure within an independent research organisation (commercial or not-for-profit) was seen to be advantageous in terms of exploiting existing expertise, but potentially disadvantageous in terms of distance from policy-making. Potential conflicts of interest in relation to policy-evaluation contracts were also seen as a problem.

Table 1: Location/types of infrastructure and associated benefits/risks

Location/type of infrastructure	Benefits	Risks
Located within an academic institution as free-standing research institute	<ul style="list-style-type: none"> - Existing pool of expertise - Ability to champion and promote high quality evaluation - Would ensure support, access, impartiality and longevity 	<ul style="list-style-type: none"> - Poor management - Naivety about political environment - Theoretical rather than practical emphasis - Purist approach stifling innovation
Located in private/not-for-profit sector as free-standing research institute	<ul style="list-style-type: none"> - Existing pool of expertise - Would ensure support, access, impartiality and longevity - Ability to champion and promote high quality evaluation 	<ul style="list-style-type: none"> - Distance from policy-making - Conflict of interest in relation to policy evaluation contracts
Located within a UK government department	<ul style="list-style-type: none"> - Proximity to spending capability - Understanding of policy and decision making - Existing pool of expertise 	<ul style="list-style-type: none"> - Open to manipulation - Unsympathetic to external agendas - Less flexible and open - Underutilisation
Located within a government-funded body	<ul style="list-style-type: none"> - Understanding of government's needs/pressures whilst maintaining independence from government - Multi-disciplinary input combining research disciplines and policy-making communities 	<ul style="list-style-type: none"> - Danger of antagonism to government - Research and policy communities fail to collaborate - A 'one-size fits all' approach for complex and diverse evaluating public policy initiatives
Virtual infrastructure, network/community of practice	<ul style="list-style-type: none"> - More feasible than a fixed infrastructure in terms of both funding and viability 	<ul style="list-style-type: none"> - Could be ignored, sidelined and under-resourced - Lack of clout compared to funded, staffed organisation

A mixed approach

Finally there was some support for a unit at arms length of government, which might be part-funded by government. It was envisaged that such a unit would be staffed both by government researchers on secondment and experts from independent or academic institutions. The advantages of such an approach were seen to be a combination of those attached to other options, particularly the utilisation of existing expertise and understanding of policy-making. It was felt that it could harness multi-disciplinary input from a variety of research disciplines and policy-making communities. On the other hand, some concern was expressed about its ability to attract staff and possible conflicts of interest.

There were some risks attached to this approach. First, it was argued that it could become antagonistic to government research considered to be sub-standard, which would in turn cause government to dissociate itself. Second, some scepticism was expressed about the ability of policy and research communities to collaborate. Third, it was suggested that the role of NIHCE provided an illustration of the expense and difficulty associated with a 'one size fits all approach' for evaluating complex and diverse public policy initiatives.

The components of infrastructure

Respondents tended not to offer detailed ideas about the precise components of infrastructure. When views were expressed the model that usually emerged was a 'hub and spokes' design, with a central unit that controlled operations and directed people to individual experts employed either within or beyond the unit, e.g. in other specialist centres or government.

There was some discussion of a virtual network model that was seen to be much less resource intensive and cheaper than the physical alternatives. However the general view was that a virtual network would be insufficient to meet the needs identified, at least on its own, and could be ignored, sidelined or under-resourced. It was argued that infrastructure needed to have a physical location to give it credibility and clout, and so that people would know there was somewhere tangible to go to for advice. It was also argued that a virtual network would also not be able to provide training, which was seen an important function of any support infrastructure. However, there was some support for having a virtual network to accompany physical infrastructure, which would facilitate information sharing.

Funding and resourcing

Respondents did not offer particular funding models, though both government itself and the ESRC were mentioned as possible sources of funding. However, a view amongst user respondents was that individual departments would be opposed to having to contribute to the core funding of any service. There was some limited indication that users would be happy to pay for any advice or support they received on an ad hoc basis.

The key to staffing infrastructure was felt to be a combination of QEPO expertise and understanding of the policy-making environment. One view was that this could be achieved by employing researchers with extensive experience of evaluation within government. A more common view was that infrastructure would need a mix of government researchers and people from the wider research community. There was some indication that researchers from established independent research institutes and units were perceived as more policy aware, though this was not a universal view.

A common view was that government researchers should be seconded to a unit, possibly on a rota basis so that every two years they could go back to their departments to refresh their skills. There was also some discussion of people from the wider research community being seconded in, though more typically it was suggested that such people should be full-time permanent employees of any unit. One important ingredient, that was identified during early discussion of these findings with the project board, is the need for input from analysts with other specialisms including economists, statisticians and operational researchers.

Some emphasis was placed on the importance of employing a heavyweight and proactive figurehead to front a unit, for example someone from within government who had credibility in both the policy and research communities. She/he would be needed to provide leadership and entrepreneurial skills and would need the charisma and drive to promote the cause of QEPO methods.

The role of infrastructure

Respondents discussed a variety of potential support roles. These included training for government researchers on QEPO methods (for which there was an overwhelming support); advice and support for government researchers around commissioning and using QEPO methods (both general and technical advice, for example statistical expertise and setting up trials); support in relation to ethical concerns and approval; and championing the use of QEPO methods in government research (although there were concerns that this should not be done at the expense of other methods).

There were mixed views about whether the role should extend to addressing the knowledge and attitudes of policy-makers towards QEPO methods, despite the fact that this was often identified as a key barrier. One view was that training courses should be run for policy-makers in the same way as government researchers.¹² A contrasting view was that 'a little knowledge might be a dangerous thing' in the hands of policy-makers and the result might be poorly commissioned research.

Two potential roles were generally ruled out. First, it was suggested that infrastructure should not have a regulatory role and that support should be voluntarily sought and be provided without strings. It was felt that regulation would inhibit use, particularly from sceptics, with the result that those most in need of support would not access it. Second, the idea of a provider unit actually conducting policy evaluation was rejected. Though the reasons for this were not always explicitly expressed, one concern was that this would lead to a conflict of interests in relation to research contracts and another was that it could be anti-competitive.

¹² Some training for policy staff is currently provided by the National School for Government, although it is not mandatory and has had limited take-up.

4 Implications of the study

Developing QEPO methods

This study clearly indicates that there is value in encouraging and enabling wider use of QEPO. However, two important qualifications need to be made. First the efficacy and the ethical basis of QEPO techniques, particularly RCTs, are contested. Even their supporters acknowledge this. Second, it is generally accepted that, like other methods, QEPO techniques are not suited to all policy contexts or questions. Thus, where they are used, it may be necessary or desirable to use them in combination with other approaches.

The authors have identified three key considerations which follow from these qualifications and which offer an opportunity to build consensus about the development of QEPO methods. First, the emphasis should be on optimising rather than maximising QEPO use. An overzealous approach is likely to be perceived as a blunt instrument, which would undermine impact. Second, the approach should be underpinned by the development of the appropriate use of QEPO methods, sensitive to policy context and, where appropriate, in combination with other methods. Third, at the point of implementation there is an important ‘triage’ role to be performed. The authors see this as a process by which project needs are analysed at the outset to review the evaluative methods being proposed and the possible role of QEPO methods within them.

If QEPO is to develop along these lines, a proactive approach is needed. This study provides evidence of interest and in some cases considerable enthusiasm, but little evidence of momentum or innovation. Despite the developments which have taken place in some quarters, the current situation could be characterised as static at best. This study has confirmed the perception of multiple barriers and disincentives, both real and imagined, such as time, cost and lack of expertise, though it is questionable whether many of these relate only to QEPO methods. Given the twin and clearly linked pillars of limited provider expertise and lack of UK demand, there are no obvious harbingers of natural change.

The study also confirmed that there is support for developing infrastructure to provide the impetus that is needed and provide support in relation to the use of QEPO methods. However, it is important to note that a range of views was expressed which voiced both ambivalence and opposition to developing infrastructure. The authors recognise the need to acknowledge and where possible respond to those arguments when moving towards the creation of any new policy evaluation service. They have sought to do this in formulating recommendations about the approach and design of a new service. In particular, the recommendations acknowledge that optimising the use of QEPO methods requires an approach based on open discussion and methodological pluralism. They also recognise that a crucial role for infrastructure will be to stimulate awareness, demand and capacity.

Recommendations

- Steps should be taken to optimise the use of experimental and quasi-experimental methods in policy evaluation.
- Optimisation of the use of QEPO should be sensitive to policy context and involve a combination of methods where appropriate.
- Project needs should be analysed at the outset in order to review the proposed evaluation methods and the role of QEPO.

Supply and demand

The authors’ reading of the study evidence is that developing QEPO requires action on two interdependent and equally important fronts: the demand side and the supply side. The demand side involves a range of key stakeholders including ministers, senior policy officials, and the Treasury as a driver of efficiency and purse-

holder in relation to discretionary spending. Capability reviews and regulatory impact assessments may provide further levers, though the latter, in particular, would need to be used carefully. Encouraging demand will require raising the profile of QEPO in a way that clarifies options in broad and simple terms. The subtle approach described in 4.1, based on optimisation and the appropriate use of methods, is likely to be more fruitful than a blanket requirement.

There are myths to be unpacked and dispelled. As has been indicated, many of the perceived disadvantages and barriers in relation to QEPO referred to in this study could equally apply to other methods. For example, the ‘too late’ approach identified by respondents underlines the need to encourage early planning of evaluation in the policy development cycle. The authors also agree with the suggestion made by one respondent that some of the ethical objections referred to in this study might more accurately be described as ‘ethico-political’. This does not make them less real, but it does suggest that a more sophisticated approach needs to be taken to dealing with them in order to build consensus and shift the debate towards best practice.

On the supply side it is clear that research commissioners are key and need to be better informed and more confident about QEPO methods. Without access to support, advice and training this is unlikely to happen consistently. Providers of research services, on the other hand need to be encouraged to tool up for increased demand. Though the authors would agree with the sentiment expressed by some respondents that ultimately this would be a market driven process, it is clear that it will need to be facilitated in the short term. Providers need a clear signal that investment will pay off. If supply is to precede growth, investment will be needed.

The authors believe that initially that may best be achieved by facilitating consortia in order to enable the limited and scattered resources referred to by respondents to be utilised and co-ordinated. A key question here is whether a dedicated provider unit (i.e. one that runs evaluation projects) is needed. The authors agree with the consensus amongst respondents in this study that it is not. They think that a dedicated provider unit could inhibit the development of expertise across the widest possible range of research providers and stifle methodological learning and innovation. Removing the provider function also reduces the risk of conflicts of interest arising in relation to policy evaluation contracts and promotes competition between providers.

The evidence from this study does underline the importance of collaboration between the demand and supply sides to move the situation forward. More effective and early dialogue between research commissioners and policy colleagues represents a key starting point. However, the authors would go further and advocate the need to foster a shared understanding and dialogue between policy-makers, research commissioners, research providers and practitioners about three central issues. These are: the place of QEPO within policy evaluation, the roles of prospective and retrospective QEPO and the functions of different types of QEPO technique.

Recommendations

- Steps should be taken to foster dialogue and understanding about the use of QEPO between policy-makers, research commissioners, research providers and practitioners.
- Dialogue should focus on the place of QEPO within policy evaluation, the roles of prospective and retrospective QEPO and the functions of different types of QEPO technique.
- The demand for QEPO should be stimulated by providing clear advice for ministers, senior policy officials and the Treasury about the role of experimental and quasi-experimental methods within the spectrum of available evaluation options.
- The supply of QEPO should be facilitated by ensuring that research commissioners are properly informed through the provision of support, advice and training, and that research providers are encouraged to invest in the development of skills, where appropriate through consortia arrangements.

Defining support needs and criteria

The collaboration suggested above is clearly ambitious and begs the question of what kind of support is needed to enable it to flourish. The evidence points to the importance of facilitation through information, advice and training. Whilst the focus of this study has been on QEPO methods, it also suggests that in order to attract users and be influential, support needs to be provided in a broader evaluative context. This may involve acknowledging that QEPO methods are not appropriate on occasions. This is consistent with the triage role identified earlier, though it is important that this role is optional so that those who have specific needs in relation to QEPO are able to access support immediately if they choose. A flexible approach would not preclude reference back to the triage stage if advisers are knowledgeable and adept at providing advice. The triage role would thus provide two levels of support, first what methods are appropriate, and second, if QEPO methods are appropriate, what the technical implications are. Such a role will provide an important signal that the infrastructure is not unduly biased in one direction or another.

A key challenge will be to balance breadth in relation to all policy evaluation methods with depth in relation to QEPO. This will mean addressing the reported skills deficit in QEPO methods within a problem/question-driven context rather than a technique-driven context where some methods are seen as being the 'gold standard'. A policy evaluation service that is able and willing to say that what is needed is a non QEPO approach is likely to have greater credibility. So qualitative expertise and the ability to direct people to alternative sources of support will be important.

Functions of a policy evaluation service

Before considering the locational, structural and funding options for QEPO infrastructure, it is necessary to define what its key functions should be and the criteria for its success. The authors have considered the functional suggestions made by respondents and suggest that a multiplicity of functions will be necessary to deliver the support needed to create impetus for the optimisation of QEPO methods. These fall within four broad categories as shown in Figure 2 overleaf.

Measuring success

By what criteria can the success of any such infrastructure be judged? Of critical importance is the need for a service to achieve practical change, as opposed to theoretical debate. This is not to say that there is no role for academic papers, but to be effective, messages will need to be practical in emphasis, and targeted and disseminated to a variety of audiences. Given the contested nature of QEPO terrain, it is vital that infrastructure is credible and trusted by a wide range of stakeholders, including those who are sceptical and uninformed. It is they, rather than merely the converted, to which infrastructure must reach out by engaging in dialogue across traditional paradigm turfs.

The evidence from this study emphasises that working with, and achieving credibility amongst, policy-makers as well as researchers, will be a vital ingredient. To do so it must be prepared to recognise and deal with the controversial status of RCTs in particular and, through engagement and dialogue, secure high level sign-up to, and active support for, the use of experimental methods where appropriate. If ethico-political arguments are to be overcome, one possibility is that advisors may be able to identify interventions suitable for experimental approaches in policy areas where there is greater political consensus or less sensitivity. Early and proactive discussion may provide opportunities to demonstrate that ethical concerns can be properly addressed, and to share good practice and experience.

Figure 2. Functions of a policy evaluation service

<i>Promotion and training</i>	<ul style="list-style-type: none"> ● Championing and promoting appropriate use of QEPO, raise its profile amongst policy-makers and researchers and stimulate demand ● Facilitating training and capacity building for a range of stakeholders at all levels, including commissioners, providers and policy-makers through both short-term measures and long-term focused intervention
<i>Knowledge and information</i>	<ul style="list-style-type: none"> ● Facilitating the build-up of departmental expertise and resources (which are clearly variable) using knowledge transfer ● Acting as a clearing house for information about methods or applications by collating and disseminating existing knowledge, and about sources of support and available resources both in the UK and beyond
<i>Advice and support</i>	<ul style="list-style-type: none"> ● Delivering an advice function which should be a multiple role focusing on when QEPO is appropriate, what methods and techniques can be used and how to implement and manage QEPO projects ● Providing practical guidance and help with the implementation of QEPO projects, statistical or analytical problems, and reporting and presenting findings in the political arena and the real world
<i>Development and innovation</i>	<ul style="list-style-type: none"> ● Facilitating greater use of secondary data and meta-analysis in order to better utilise data and extend available knowledge about methodology ● Enhancing standards and providing quality assurance through the collection and dissemination of good practice and the provision of independent reviews (The authors agree with the consensus amongst respondents that it should not act as a regulator) ● Clarifying and codifying QEPO approaches and help to develop QEPO technology and facilitate leading edge development

Charging

Finally there is the question of charging. Though there was some indication of a willingness to pay for advice, the evidence suggests that immediate advice needs to be free at the point of delivery since charging for it would act as a deterrent to those who might be persuaded to seek it. Moreover, there is an argument that an effective free triage service will be cost-effective for departments because it will enhance the efficacy of research commissioning. For infrastructure to be viable in the longer term the authors think that other functions such as training and technical advice and support once a project is live would need to be charged out.

Recommendations

- Support for the development of QEPO methods needs to be provided through the provision of information, advice and training in a broad evaluative context.
- The functions of a policy evaluation service should focus on four key areas: promotion and training, knowledge and information, advice and support, and development an innovation.
- The credibility of any policy evaluation service will need to be built carefully through the dissemination of practical messages and by addressing perceived ethical concerns about QEPO methods.

- The provision of initial advice on QEPO methods by a policy evaluation service should be free, but a charging system for ongoing project support should be introduced.

Determining the location and structure

Now comes the question of how QEPO support should be provided. Here the authors have had to make sense of conflicting opinions about the benefits and risks of various options. Some perceived risks and benefits were attributed to more than one option. On balance, despite the fact that the dominant view amongst respondents was for a structure within government (probably alongside GSRU), the authors' view is that the best option would be an external location. They have arrived at this view because it is clear that wherever it is located, infrastructure will need to foster an exchange of ideas and experience, including through secondment of personnel. The authors believe that that a structure outside and independent of government is more likely to facilitate this and maximise involvement, openness and creativity, albeit that the issue of confidentiality and policy sensitivity will need to be addressed. This could be either an academic or an independent research provider and the key organisational specification criteria should be two-fold. First, the organisation will need to have a broad range of existing methodological and substantive expertise to enable capacity and ensure credibility across the policy and research communities. Secondly, an organisation is needed that is demonstrably committed to the development of QEPO methods, both intellectually and financially. Whichever organisation delivers a policy evaluation service, strong leadership and entrepreneurial skills will be essential ingredients for its success.

It is recognised that no single research institution will be expert in all methods however. In order to ensure that it is strongly integrated into the widest possible network of methodological expertise, the authors would advocate that any new policy evaluation service should establish clear links with the ESRC's National Centre for Research Methods and its node network. Whilst the authors do not think it would be appropriate for it to be another node, they do think that some arrangement whereby it has affiliated status would be potentially highly beneficial. These arrangements could be consolidated through creating a role for the NCRM director in its governance structures. The authors recognise that the NCRM already has established links with the government social research network and with a wide range of methodological, training and research ethics activities, which could contribute to the viability and early success of a new policy evaluation service.

The authors have also carefully considered the combination of physical and virtual structures discussed by respondents. They believe that these options are not mutually exclusive and that both will be needed to ensure the widest possible engagement and dissemination of good practice. They therefore envisage the infrastructure as a system rather than a single unit, with a hub and spokes type as envisaged by some respondents, with a physical unit at the core, part of whose role would be to develop and support virtual networking opportunities.

Recommendations

A policy evaluation service should:

- be independent and located outside government in order to maximise involvement, with appropriate arrangements for confidentiality;
- have strong links to the ESRC's National Centre for Research Methods and its node network;
- be provided from within an academic or independent research organisation with a clear commitment to developing the use of QEPO methods and a broad range of methodological and substantive expertise;
- be underpinned by strong leadership, entrepreneurial skills and the secondment of personnel to maximise opportunities for exchanging ideas and expertise;
- include both physical and virtual infrastructure.

5 Conclusions

This study suggests that both inside and outside government there is support for widening the use of QEPO methods. This view is not universal and, of course, strong epistemological differences do exist in the wider research community and will continue to do so. However, these should not preclude proper consideration of the use of QEPO methods and their use in policy evaluation where appropriate.

Whilst disagreements will remain, there is scope to build greater consensus about the role of experimentation within policy evaluation, including on key ethical and practical issues. As part of this process, there is a need to de-mystify and even de-stigmatise some such methods. Key to this will be collaboration on the demand and supply sides within and outside government, and the creation of an agenda, which is premised on the optimum use of experimentation.

The barriers identified in this study underline the need for intellectual, practical and financial investment in the development of QEPO methods. Given the current lack of momentum, in the short and medium term, such investment will require facilitation. This provides a strong case for the development of infrastructure to provide a policy evaluation service based on shared dialogue and collaboration between key stakeholders.

The location of infrastructure is a key overarching consideration. It is argued here that on balance a policy evaluation service would best be delivered outside government, but drawing heavily on government expertise and policy know-how. Ultimately, factors other than location will determine success. Adequate financial investment will be a key issue and a range of other factors will be significant. First, a proactive and entrepreneurial approach will be needed to drive the QEPO agenda forward and establish the service's credibility across institutional and methodological boundaries. Second, the provision of a confidential service will be vital to address concerns about sharing sensitive information outside government. Third, advice and support should initially be free at the point of delivery so as not to deter those sceptical about the use of QEPO methods.

References

Campbell, S., Benita, S., Coates, E., and Davies, P. *Analysis for policy: evidence-based policy in practice* London: Government Social Research Unit

Chapman, B., Mirrlees-Black, C. and Brawn, C. (2002) *Improving public attitudes to the Criminal Justice System: The impact of information*. London: Home Office. Home Office Research Study 245.

Greenberg, D. and Poole, L. (2007) *Designing a social experiment for the UK: how it was done and some lessons learned*, Evidence and Policy 3:1:5-29

Hale, J., Roberts, C., Moore, L. and Conway, R. (2006) *Primary School Free Breakfast Initiative in Wales: application of a cluster randomised trial in government research* (Unpublished paper)

Hall, N., Hoggart, L., Marsh, A., Phillips, J., Ray, K., and Vegeris, S. (2005) *The Employment Retention and Advancement scheme – the early months of implementation: Summary and conclusions*. London: Department for Work and Pensions. Research Report No 265.

Jowell, R. (2003) *Trying It Out: The Role of Pilots in Policy-Making* London: Cabinet Office.

Leston, J. and Watmough, M. (2005) *Providing pensions information and advice in the workplace where there is little or no employer contribution:*

Pilot evaluation findings based on survey and qualitative research. London: Department for Work and Pensions Research Report No 294.

Miers, D., Maguire, M., Goldie, S., Sharpe, K., Hale, C., Netten, A., Uglow, S., Doolin, K., Hallam, A., Enterkin, J. and Newburn, T. (2001) *An Exploratory Evaluation of Restorative Justice Schemes*. London: Home Office. Crime Reduction Research Series Paper 9.

Uglow, S., Doolin, K., Hallam, A., Enterkin, J. and Newburn, T. (2001) *An Exploratory Evaluation of Restorative Justice Schemes*. London: Home Office. Crime Reduction Research Series Paper 9.

Natarajan, L., Smethurst, L., Stratford, N., and Taylor, R. (2005) 'The implementation of a randomised controlled trial: improving the measurement of impact in the Job Retention and Rehabilitation Pilot using a screening tool developed for the DWP, UK' in *Best Practice in the evaluation of labour market programmes for vulnerable groups* Nicaise, I. (ed) Hoger Instituut voor de arbeid, Leuven.

Shapland, J., Atkinson, A., Colledge, E., Dignan, J., Howes, M., Johnstone, J., Pennant, R., Robinson, G. and Sorsby, A. (2004) *Implementing restorative justice schemes (Crime Reduction Programme): a report on the first year*. Home Office Online Report 32/04. London: Home Office (at <http://www.homeoffice.gov.uk/rds/pdfs04/rdsolr3204.pdf>).

Shapland, J., Atkinson, A., Atkinson, H., Chapman, B., Colledge, E., Dignan, J., Howes, M., Johnstone, J., Robinson, G. and Sorsby, A. (2006) *Restorative justice in practice – findings from the second phase of the evaluation of three schemes*. Home Office Research Findings 274. London: Home Office (at <http://www.homeoffice.gov.uk/rds/pdfs06/r274.pdf>)

Shapland, J., Atkinson, A., Atkinson, H., Chapman, B., Dignan, J., Howes, M., Johnstone, J., Robinson, G. and Sorsby, A. (2007) *Restorative Justice: the views of victims and offenders – the third report from the evaluation of three schemes*. London: Ministry of Justice Research Series 03/2007
(at <http://justice.gsi.gov.uk/www.justice.gov.uk/publications/research.htm>)

Starkey, F., Moore, L., Campbell, R., Sidaway, M. and Bloor, M. (2005) *Rationale, design and conduct of a comprehensive evaluation of a school-based peer-led anti-smoking intervention in the UK: the ASSIST cluster randomised trial* [ISRCTN55572965]. *BMC Public Health* 2005, 5:43. 22 April 2005.
<http://www.biomedcentral.com/1471-2458/5/43>

Tapper, K., Murphy, S., Moore, L., Lynch, R. and Clark, R. (2007) *Evaluating the free school breakfast initiative in Wales: methodological issues*. *British Food Journal*. 109:3:206-215.

Ministry of Justice Research Series 8/07

There has been increasing interest in improving the efficacy of policy evaluation. This report looks at the arguments concerning the development of infrastructure to support the use of experimental and quasi-experimental methods in policy outcome evaluation. It sets out proposals for the creation of a new policy evaluation service.

ISBN 978 1 84099 088 1

© Crown copyright 2007

Produced by Ministry of Justice

Point of contact research@justice.gsi.gov.uk