King's College LONDON

Cillege

University of London

psi

Policy Studies Institute

Assessing the impact of research on policy:

A review of the literature for a project on bridging research and policy through outcome evaluation

Final report, February 2008

Annette Boaz, Siobhan Fitzpatrick and Ben Shaw

King's College London

King's College London is one of the top 25 universities in the world (Times Higher 2007) and the fourth oldest in England. A research-led university based in the heart of London, King's has 19,700 students from more than 140 countries, and 5,400 employees. King's has an outstanding reputation for providing world-class teaching and cutting-edge research. The College is in the top group of UK universities for research earnings and has an annual income of approximately £400 million. An investment of £500 million has been made in the redevelopment of its estate. King's has a particularly distinguished reputation in the humanities, law, engineering, and social, health and natural sciences.

Policy Studies Institute

The Policy Studies Institute (PSI) is one of the UK's leading independent research institutes, highly regarded for rigorous and impartial evaluation of policy in the UK and Europe. PSI's Environment Group seeks to undertake interdisciplinary, policy-relevant research of the highest quality, which recognises the interaction at every level between the environment, the economy and society generally. Environmental research at PSI employs a wide range of research methods, from formal modelling and scenario building to various styles of policy analysis and evaluation.

<u>Contents</u>

Title: **Page:**

This is a condensed version of the report; full references and appendices are available in an additional document.

<u>Literature Review</u>

1. Background to project

There is a growing interest in evaluating the impact of research. This interest can be discerned in UK research councils (10,15,38[1]), international organisations (3,37,145), the European Union (46,55,56) and academia (45,108,139). Furthermore, it can be seen in a wide range of different disciplinary and policy domains. Recent evaluations have considered both the broad impacts of research and its specific impacts on policy. With these developments has come an interest in a wider variety of measures of impact, reaching out to assess the non-academic influence of research investments (100,120,144).

The factors motivating organisations and individuals to consider research impact are manifold. They include accountability, performance, promotion of organisational achievements, learning and moderating between competing stakeholders or interests (10, 14). The pressure for accountability is particularly clear in a number of recent documents (15,103). Increased government investment in research and development in the UK (69) has been accompanied by a new programme of independent reviews to evaluate the quality and use of science in government departments.

In recent years Defra has invested in better understanding the role of evidence in policy making, with a focus on the demand for evidence-based policy (Defra, 2006), the use of existing research within the department (Taig, 2004), the need to monitor and evaluate policy (Defra, 2007) and the potential for bridging the gap between research and policy through collaboration and advisory committees (Ashridge, 2006). However, analysis within and for Defra has primarily focused on the process of managing the evidence base. It does not refer specifically to evaluating research impact on policy making, other than to conclude that assessing the success of science policy interaction is important but challenging (Scott et al, 2005). A recent strategy document concluded: 'In order to strengthen the evaluative aspects of the programme and to measure our progress to date in commissioning and using research to support policy, we have tendered an outcome evaluation...the project will explore good

---

[1] Each paper included in the review has been given a number to avoid long references in the text. The numbers can be used to find both the references and the data extraction sheets for each study.

practices drawn from international experiences in evaluating impacts of research on attaining policy outcomes. It will apply these experiences to the wider sustainable consumption and production and waste and resources evidence programmes.' (Defra, 2007:31)

The Centre for Evidence & Policy at King's College London and the Environment Group at Policy Studies Institute were commissioned by Defra to conduct a literature review to explore and examine methods for evaluating the impact of research on policy outcomes that might be appropriate to the Sustainable Waste and Resource Management (SWRM) and the Sustainable Consumption and Production (SCP) research programmes. The review is part of a larger study conducted in partnership with Technopolis and GHK Consulting that will lead to a new evaluation framework for assessing the impact of Defra-funded research on policy outcomes. The literature review is designed to ensure that the development of the evaluation framework is informed by the best available evidence.

The review aimed to answer the following question:

How do you (best) evaluate the impact of research programmes on policy?

To answer this question, the following sub-questions were addressed:
- How do people evaluate the impact of research programmes?
- What are the advantages and disadvantages of different approaches?
- Which methods are most effective?
- Which methods offer value for money?
- Are there any specific issues to consider when evaluating the impact of research on environmental policy?

The review questions reflect the priorities expressed by Defra in the original invitation to tender document.

This report begins with a brief description of the methods used to conduct the review. The next section describes a mapping exercise that categorises the literature around a

number of key variables. While mapping is a relatively descriptive process, it generates insights into the research landscape of research impact evaluation. The next two sections outline and discuss what can be learnt from the literature with regard to the review sub-questions, and outline gaps in the literature relevant to these questions. The final section returns to the main review question: How do you (best) evaluate the impact of research on policy?

2. Methods

The review focused on English language publications covering a period of twenty years (1987-2007) and sought to identify:

- Evaluations of the impact of research on policy
- Reflective pieces about evaluation of the impact of research on policy
- Descriptions of methods for evaluating the impact of research on policy
- Reviews of methods for more general impact evaluations
- Evaluations of the effectiveness/value for money of different approaches

Particular effort was made to identify papers from other countries and papers with specific reference or relevance to environmental, waste and sustainable consumption and production policy. However, the majority of the papers found through these additional searches were related to agriculture policy, and those that did refer specifically to SCP or SWRM did not meet the inclusion criteria for the study, largely because they did not address any of the review questions.

The search strategy included four elements: database searches, web searches, citation tracking and expert contacts. The sources comprised ten key databases, three library catalogues and thirty relevant organisations. Citation tracking involved following up references and contacts emerging from the documents identified through the initial stages of the search, and proved to be a particularly fruitful source of relevant material. The final part of the search aimed to ensure that the review captured the most up-to-date knowledge on impact assessment. This was achieved through holding sessions with the broader study team responsible for the evaluation, and through asking expert contacts for feedback on the draft report and for their advice regarding the inclusion of information from any unpublished papers or work in progress.

A total of 351 papers were identified through the search, of which 156 met the selection criteria for inclusion in the review. A data extraction sheet was completed for each included study, and these are attached as Appendix A.

A detailed account of the review methodology can be found in Appendix B.

3. Mapping the literature

The first stage of the analysis involved mapping and categorising the literature around a number of key variables: topic, country of origin, the way in which the papers and reports address the review question, and the methodologies used or advocated for conducting impact evaluations of research and development on policy. The data tables for the figures one to four are attached in Appendix C. Figures 5 and 6 are derived from the methods table (Appendix D).

Figure 1: Method of Identification
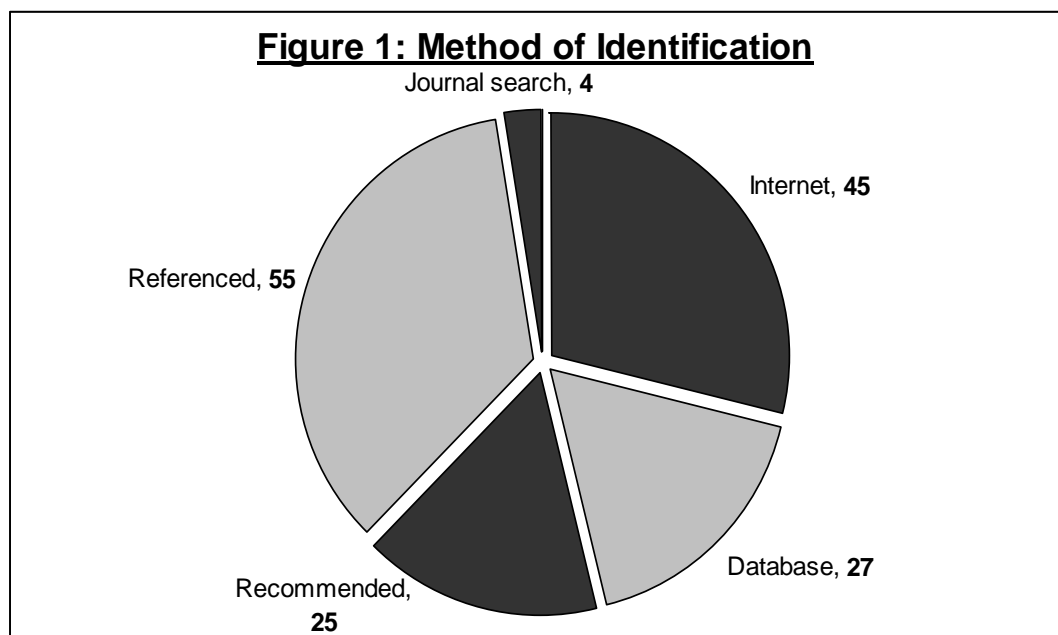
The majority of the papers were identified through the internet search and from the reference lists in papers and reports. Personal recommendations also made a strong contribution. Although only 27 papers were identified through the database searches, this approach did provide access to papers not identified through other sources, reinforcing the value of a mixed method search strategy.

Figure 2: Paper includes:

Evaluation of effectiveness/ value for money of methods, **13**

Other, **15**

A review of methods for evaluating impact more generally, **24**

An evaluation of the impact of research on policy, **58**

A description of method (s) for evaluating impact of research on policy, **42**

A reflective piece about evaluation of the impact of research on policy, **74**
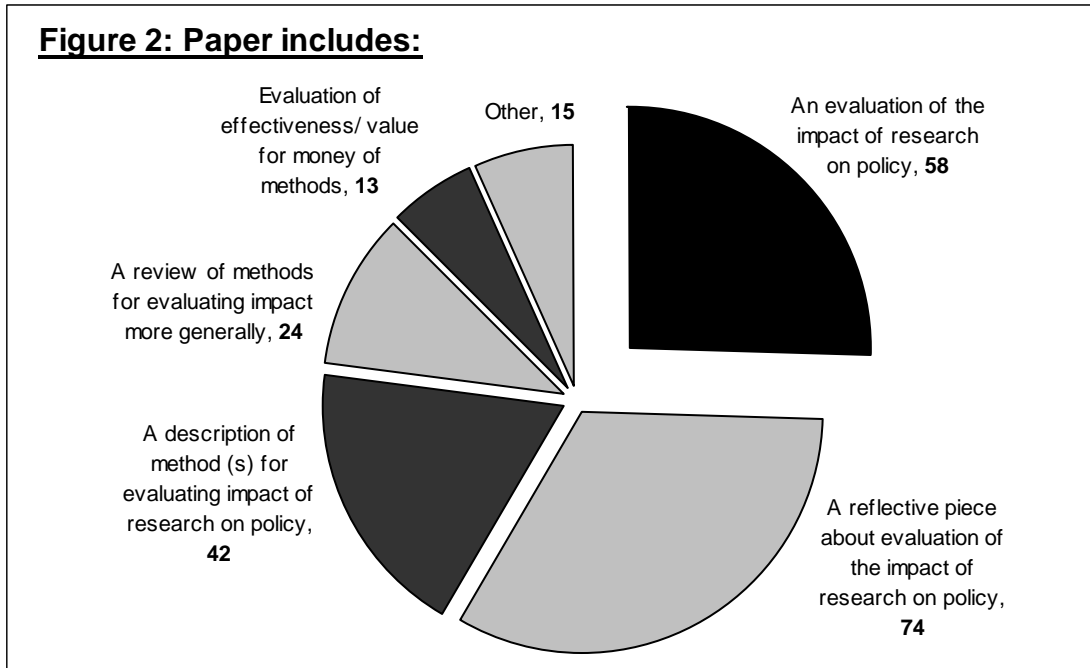
This review particularly focused on evaluations of the impact of research on policy. Of the 156 studies, 58 reported on empirical studies of the relationship between research and policy, of which 42 were specific evaluations of the impact of research on policy. Most of the remainder were reflections on (or descriptions of) methods or approaches to impact evaluation.

Figure 3: Field

Environment, **26**

International Development, **26**

Science, **14**

Utilization literature, **13**

Policy Research, **4**

Evaluation literature, **38**
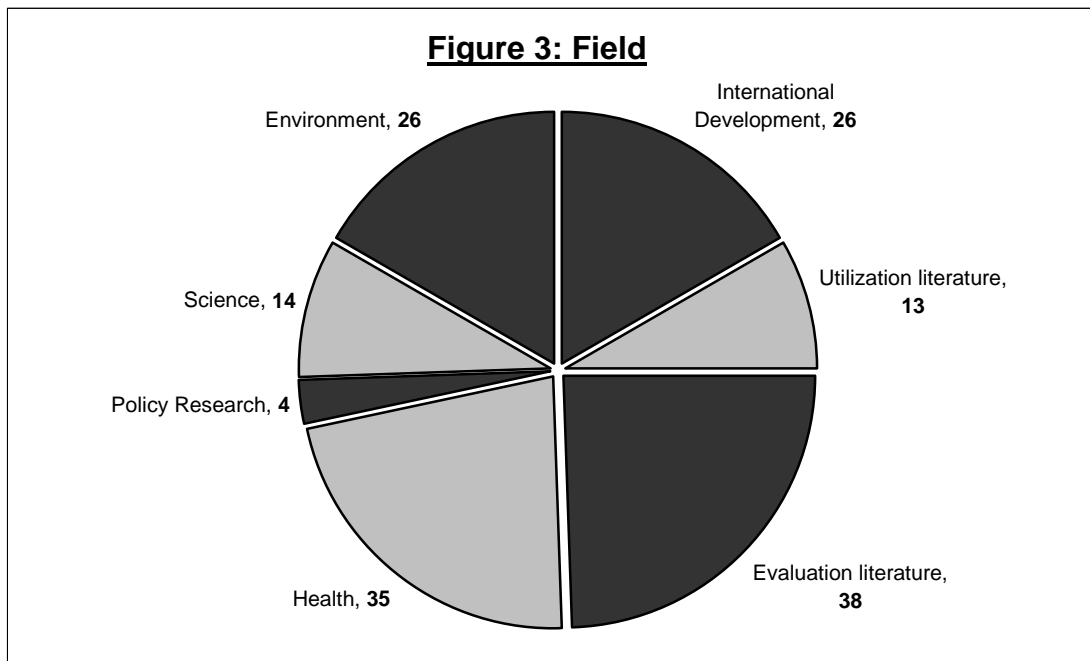
Health, **35**

The review identified a wide-ranging literature, reflecting a cross-sectoral interest in the issue of research impact on policy. In particular, there are large health and international development literatures as well as a generic literature on utilisation and evaluation. Particular effort was made to identify papers with topical relevance to Defra, resulting in the identification of 37 environmental and science papers.

Figure 4: Country of Origin

Other, 13
Australia, 8
Canada, 20
Finland, 4
Switzerland, 3
The Netherlands, 9
UK, 61
USA, 38
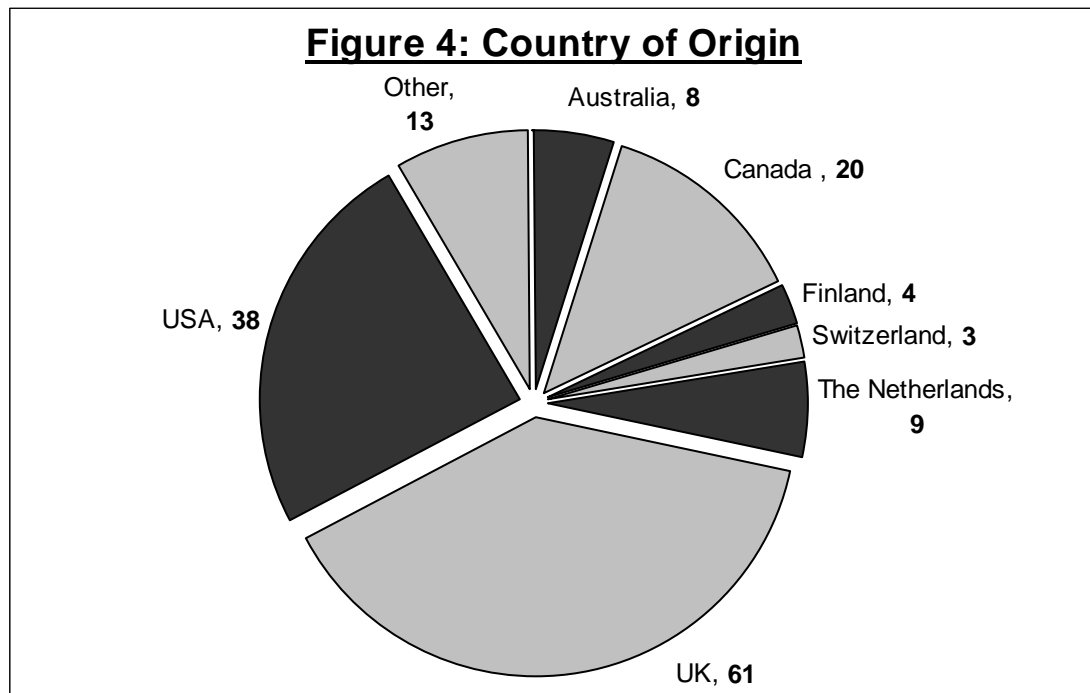
The majority of the papers come from the UK, USA and Canada, with smaller numbers from Australia, the Netherlands, Finland and Switzerland. Studies were also identified from a diverse range of 'other' countries including Brazil, Denmark, Indonesia, Norway and Spain. The spread of countries is likely to reflect, to some extent, the focus on English language publications.

Figure 5: Approaches and methodologies
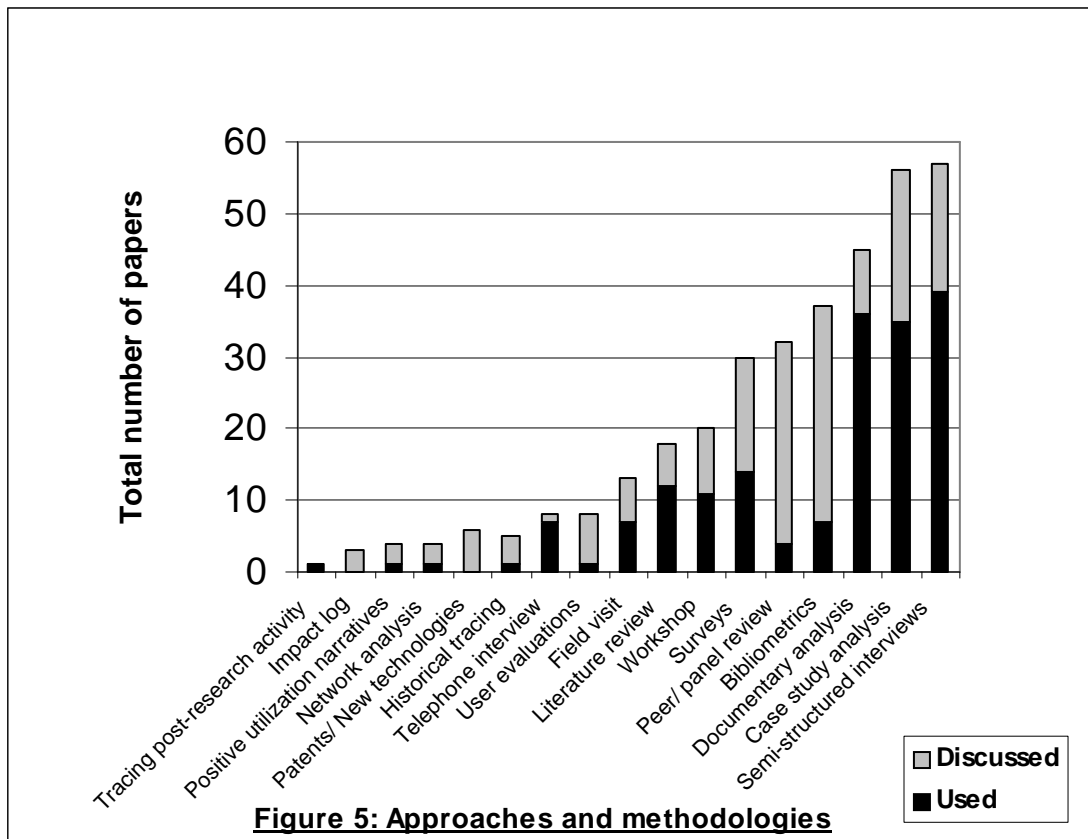
A wide range of research methods and approaches were applied to the study of research impact. The most frequently mentioned and used were semi-structured interviews, case studies and documentary analysis. This is consistent with another recent review of impact studies (65a). However, most papers mentioned more than one method. Interestingly, while bibliometric analysis was frequently cited in empirical studies, this method was rarely used to measure the impact of research on policy[2]. By contrast, telephone interviewing was used almost as frequently as it was discussed. There were many more references to, and uses of, qualitative methods than quantitative methods. A short description of each method can be found in table 1 (Appendix D).

---

[2] See Hanney at al (2005) for an interesting discussion of the use of bibliometrics as a method for assessing the impact of a set of research studies by Alberti and colleague on subsequent research and on clinical practice.

Figure 6: Frameworks for structuring and interpreting data
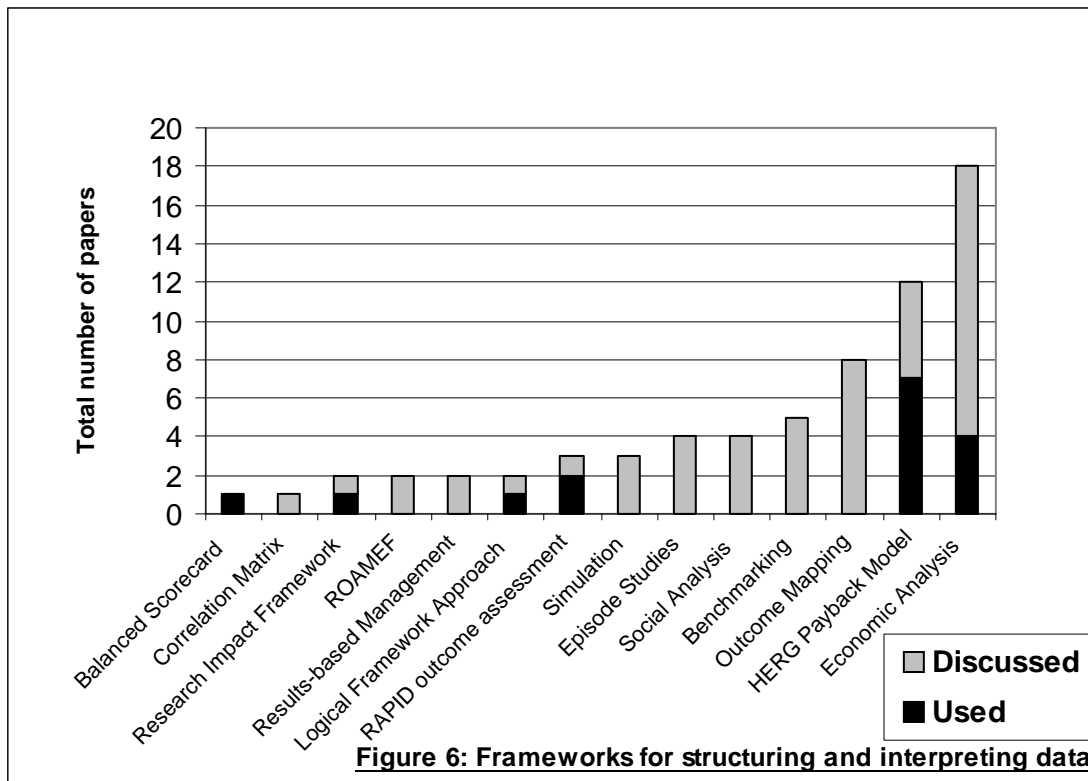
Many different interpretation frameworks are discussed in the literature but only a small number are actually used in impact evaluations. In particular, RAPID Outcome Assessment, the HERG Payback Model and economic analysis are applied in empirical studies (although the last normally forms part of a wider analysis). Descriptions of the HERG and RAPID models are attached as Appendices E and F as these provide particularly useful insights into the process of research impact evaluation. A short description of all the frameworks can be found in Table 1 (Appendix D).

While mapping is a relatively descriptive process, it generates insights into the research landscape on research impact evaluation and provides a valuable backdrop to the more detailed analysis below.

4. Using the literature to answer the review questions

This section of the report outlines and discusses what can be learnt from the literature with regard to the review sub-questions before addressing the key question: How do you (best) evaluate the impact of research on policy? The literature review has been particularly timely as it taps into a rich seam of knowledge at a time of rapid change in evaluative practice. Many of the key papers identified have been published within the last couple of years and describe attempts to develop and apply new approaches to evaluating research impact.

The relationship between knowledge and decision making has been a matter of interest for hundreds, if not thousands, of years (Shadish et al, 1991). Much can be learnt about the general relationships between knowledge and policy from the wider literature, and this has been summarised in a series of reports and papers published by the Research Unit for Research Utilisation (38,109,138,139). The diverse models of the relationship between policy and research have consequences for the evaluation of research impact. For example, it has been acknowledged (Kothari et al, 2005) that a more instrumental model of research use is closely allied with immediate use, while conceptual or enlightened use might be expected to occur in an eclectic fashion over a long period of time.

So what do we mean by impact? An evaluator may ask many different questions about a project or programme; ranging from has it worked to why and how it has (or has not) worked and whether or not it has been acceptable to participants. The particular focus of this review is the impact of research upon policy. Multiple terms are used to describe research impact, including outcomes, benefit, payback, translation, transfer, uptake and utilisation (11,25,50). The terms are often used interchangeably, although there have been some efforts to distinguish between outcomes and impacts. For example, the International Development Research Centre (IDRC)[3] argues that it is helpful to think in terms of behavioural outcomes rather than longer term policy impacts. These different terms refer to a shared interest in change that lies beyond the research process and its primary outputs. They are concerned
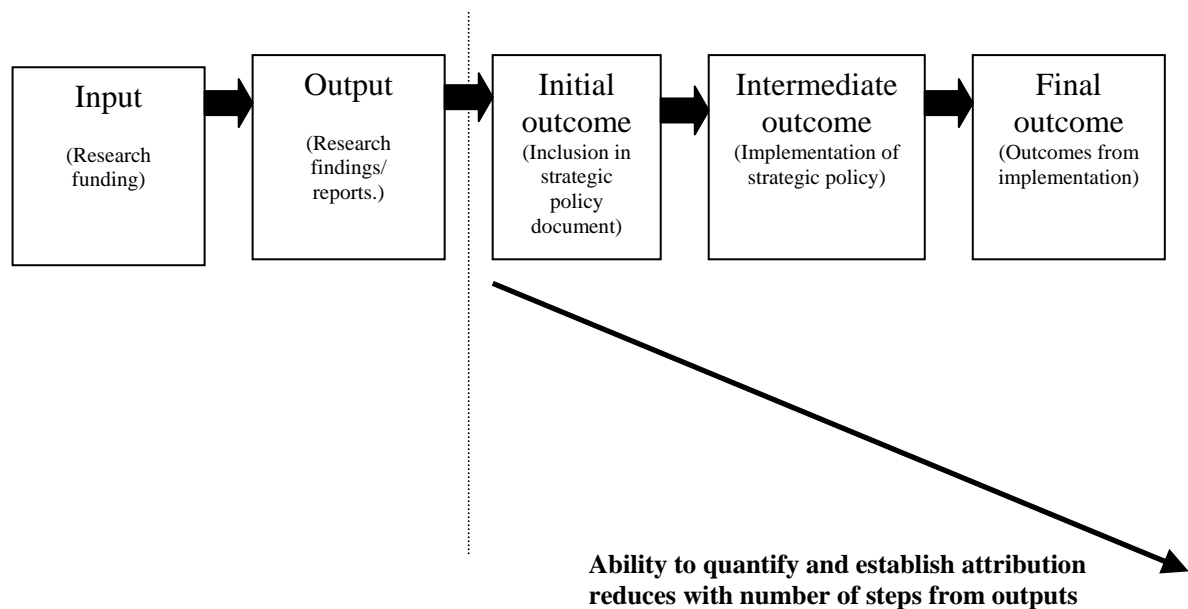
---

[3] http://www.idrc.ca/en/ev-1-201-1-DO_TOPIC.html

primarily with policy outcomes or impacts rather than outputs or process indicators. Outcomes have been defined as:

> The end results of a programme for the people it was intended to serve. (Weiss, 1998: 8)

With the interest in outcomes comes problems with attribution. The key question for evaluators is therefore the extent to which impacts can be attributed to the use of specific research outputs.

Figure 7: Looking at outcomes: losing attribution

Input	Output	Initial outcome	Intermediate outcome	Final outcome
(Research funding)	(Research findings/ reports.)	(Inclusion in strategic policy document)	(Implementation of strategic policy)	(Outcomes from implementation)

Ability to quantify and establish attribution reduces with number of steps from outputs

The process by which the research is done might also have an impact on policy. For example, research in collaborative projects may have an impact prior to the production of research outputs.

Many approaches remain primarily concerned with the quantity and quality of research outputs rather than outcomes (9,22,23). For example, bibliometrics are used to quantify research publications, while peer review panels consider the quality of the research produced by projects and programmes (often in relation to generic indicators of research quality rather than its fit to wider programme objectives) (22,92). Attribution becomes more difficult as the evaluator travels away from the initial

outputs, to look at different types of impact within the policy process. This has been described as going 'through the door' rather than 'to the door' of the policy maker (144). Only a few studies have focused on how policy is made 'behind the door' and how research might contribute to that process. One study distinguishes between the use of research conclusions and research recommendations (20).

Others have promoted a shift in language, preferring to focus on influence rather than impact. Carden (26) stresses the difficulties in ascertaining attribution as the research is often intertwined with too many other factors by the time impact has occurred. He encourages evaluators to focus on changes in behaviour as indicators of broader change. Nutley et al make a similar point:

> Research use may thus be more about transformation than straightforward application. There may, however, come a point where such refinements are so extensive that it is no longer legitimate to refer to this process as 'research use' at all. (Nutley et al, 2007: 59)

It is perhaps not surprising that the task of quantifying and tracking research outputs (using techniques such as basic bibliometrics) has often been considered to be a more deliverable and thus attractive task. Furthermore, recent developments in the practice of documenting and assessing citations have improved the potential contribution of bibliometrics. One study has attempted to track citations in policy and professional guidance and the mass media (104) and there is a call for more sophisticated analysis to move beyond, for example, zero order approximations by which all impacts are awarded the same value (144).

The evaluation process should be driven by what is meaningful in terms of outcomes, not by what is measurable. Whilst this review focuses on the difficult task of measuring outcomes, methods for evaluating processes are also included as it is clear that understanding outcomes requires an exploration of processes, barriers and facilitators as well as the end results.

4.1 How do people do it?

The literature (97,125) argues that evaluators should be driven by the need to identify methods which are the most 'fit for purpose'. This is not a straightforward task. Some methods are very flexible and lend themselves to a wide variety of evaluative tasks. For example, semi-structured interviews can be used to track forwards from a piece of research or a programme, or backwards from a policy document. They can be used with a wide range of stakeholders to explore a diversity of issues. Other methods, such as network analysis, have more specific applications. A few approaches, including cost-benefit analysis, require a particular set of technical skills in order to be successfully applied to impact evaluations (92).

The approach selected also needs to be sensitive to the context in which the evaluation will be undertaken (108,139). There are some interesting differences between the approaches to evaluating the impact of research on policy in different countries and contexts. One critical factor is the outcomes of interest. In the field of health care there has been a considerable volume of empirical work exploring research impact. However, the main focus of this work has been the impact on clinical practice rather than policy (34,45,114). By contrast, evaluations of the impact of scientific research seem to focus more directly on either changes in science policy or on the impact of research on technological innovation (e.g. through patent analysis) and also on national and regional/sub-regional economic development (28,59). To some extent, these differences reflect particular departmental responsibilities: while the Department of Health is primarily concerned with delivering services through the National Health Service, the Department of Innovation, Universities and Skills has a more facilitative role in promoting the role of science in the UK economy. A particular value of the Health Economics Research Group (HERG) payback model that has emerged from health services research in the past few years lies in its encouragement of a more systematic and transparent consideration of the outcomes of interest for each evaluation. While it does not require all evaluations to consider all dimensions, it helps to put the dimensions in context and to ensure that those selected are the ones of critical interest.

A considerable amount of impact evaluation is conducted in the field of international development, where the emphasis seems to be on qualitative, participatory evaluations

with a focus on learning and service improvement. Most of the evaluations use case studies or mixed method approaches (1,30,50,87) and participation is encouraged, primarily through workshops. While engaging potential users in the process is complex and time consuming, Carden argues that it has the potential to increase both relevance and use (26b[4]). A further distinction between evaluation in international development and other domains is the use of field visits, where evaluators observe practice on the ground (50,135). This can generate valuable insights, particularly when triangulated with data from other sources.

The reliability for policy purposes of the more qualitative data collected in both international development evaluations and other fields seems to vary. For example, storytelling has been used to explore and explain impacts (1,127). While stories can be based on research evidence, they can also be largely anecdotal and highly subjective. The focus on constructing positive utilisation narratives (1,39,72) strays dangerously close to the line between evaluation and promotion. International development evaluations pose important questions about who conducts evaluations (professional evaluators, programme staff, service users), stakeholder involvement in the evaluation and the value of a mixed method, qualitative approach.

By contrast, the US conceptual literature reflects the requirement outlined in the US Government Performance and Results Act for greater use of quantitative methods and performance indicators in impact evaluations (9). However, commentators have struggled with the applicability of modelling and economic analyses to the evaluation of research impact (132). Zilberman and Heiman (146) developed an economic formula to assess the net benefit of research that led to innovations, but acknowledged the empirical challenges of obtaining actual estimates. They suggest that the most important contribution of impact assessments of research to policy may be preventative (i.e. to screen wasteful policy proposals). The lack of evidence available (146), the quality of existing data and the uncertainties characteristic of the research process (82) make methods such as cost-benefit analysis (used so widely in industry to explore costs involved and benefits generated), extremely difficult to apply (63). Empirical evaluations conducted in the US tend to demonstrate continued reliance on

---

[4] An 'a, b, c or d' after the number denotes a paper added after the consultation with experts.

a mixed method approach to overcome the limitations of individual methodologies by a process of triangulation (80).

Panel reviews are the main method used for evaluating the impact of European Union (EU) research programmes (9). There have been a number of recent reviews exploring alternative methods for evaluating EU investments such as bibliometrics, econometrics and social analysis (48). However, the value of panels continues to be emphasised and other methods have been advocated as a complement rather than a replacement. Although panel reviews attract criticism for their reliance on experts, they do help to build ownership through the participation of key individuals from different countries within the EU. This is particularly important in the European context and underlines the importance of selecting methods that are both fit for purpose and appropriate to the needs of key stakeholders. While panel reviews are the predominant approach in the EU, other methods have been applied. One example is a large scale internet survey of researchers funded as part of the 5th Framework programme and conducted in 2004 (62).

A European Commission report on different approaches to evaluating energy research in EU countries (46) identified a shift in practice away from project management assessments (e.g. project delivered on time and to budget) towards an interest in the impacts on science, policy and industry. While the paper described diverse practices within Europe (for example, in Denmark there is a focus on evaluating the use of research within target groups) the study also identified common interests across the EU, including the need to demonstrate the additionality of research investments and a commitment to developing meaningful quantitative indicators.

Practice varies considerably between different countries and policy contexts. There are common themes regarding: the interest in developing and applying new methods, such as network analysis (5,48,80,88); new understandings of change,including the increased focus on evaluating behaviours (42); the need to develop existing methods to ensure they are fit for purpose, such as new developments in citation analysis (29,106); and the promotion of mixed method approaches. However, for evaluation to be relevant and usable in different contexts, some diversity is likely to remain in evaluation practice. For example, there is an emphasis on stakeholder engagement in

evaluation in international development. Similarly, whilst the use of peer panels in EU evaluations may be complemented in the future by a wider range of methods, they are unlikely to be replaced given their valuable role in building ownership and consensus.

The evaluation of research impact therefore calls upon a diverse range of methods including; semi-structured interviews, documentary analysis, field visits and observations, surveys, bibliometrics, patent/new technology tracking, panels and peer review, workshops and focus groups, historical tracing, positive utilisation narratives, literature review and network mapping and analysis.

4.2 What are the advantages and disadvantages of different approaches?
The wide range of different approaches used when analysing the impact of research on policy are listed in Table 1 (Appendix D). Their advantages and disadvantages are included in the table, along with a short description of each method. These can be grouped as follows:

- **Qualitative methods, including semi-structured interviews, documentary analysis, field visits and observations.** These methods are used to generate rich descriptive and explanatory data that can be used to look at both utilisation pathways and the policy context in which utilisation is or is not taking place. However, they are often time- and resource-intensive, and it is difficult to generalise their findings.

- **Quantitative methods including surveys, bibliometrics and patent/new technology tracking.** These methods are used to quantify the impact of research. They are suitable for repeat analyses and comparisons, can be used to manage large amounts of data and can be very cost-effective. However, they are often difficult to use in the analysis of research impact on policy. Survey response rates can be poor and bibliometric data usually focus on the quantity (rather than quality) of research outputs rather than outcomes.

- **Panels and peer review.** This is a relatively flexible and cost-effective approach to evaluation. Experts and peers bring status and credibility to the process and build ownership of the findings. These methods tend to be used to enhance the quality of the research rather than to assess policy impact.

Furthermore, panels are open to accusations of lack of objectivity and are heavily reliant on the quality of their membership.

- **Workshops and focus groups**. This is an interactive, consensus building approach, but is not suitable for all topics (for example sensitive topics that generate polarised views) and can lack rigour and objectivity.

- **Process tracking including historical tracing, positive utilisation narratives, tracing post-research activity and impact logs.** These methods focus on exploring the utilisation pathway. They rely heavily on the quality and relevance of existing records (or record keeping in the case of impact logs), documents and recall of participants in the process.

- **Literature review.** This method is usually used as part of a mixed method study to scope a topic and place the impact evaluation in its wider context.

- **Network mapping and analysis**. This approach seeks to construct realistic accounts of the nature of relationships between knowledge and policy. Other research methods are used to generate the data required for constructing maps of the networks that exist between individuals and organisations.

Different methods and approaches can produce very different assessments of the relative impact of research, thus reinforcing the value of mixed method approaches. The individual methods listed above are rarely used in isolation. For example, case studies usually involve a combination of qualitative and sometimes quantitative methods to explore utilisation (98,107). Most of the papers identified promote the idea of mixed method approaches to exploring research impact, whether as part of a case study or not (6,146). In particular, they argue for a combination of qualitative and quantitative approaches, acknowledging the different contributions the two approaches can make.

In addition to selecting methods, there is some question about who should be involved in conducting the evaluation. For example, user evaluations involve stakeholders in producing the study, while expert and peer panels also represent a more interactive, consensus building approach. These approaches have the advantage of building ownership and fostering individual and organisational learning. However, while these approaches can be cost-effective, there can be challenges in terms of quality,

objectivity and bias (although it is important to note that not all external reviews are objective just as not all internal reviews are subjective). In a review of methods for assessing research impact, Hovland (72) suggests that a strong evaluation would include elements of both internal and external review. The potential of a mixed approach is echoed elsewhere for practical and methodological reasons. For example, it has been argued (144) that internal evaluators would be in a better position than external evaluators to keep a record of impacts, both in the short term and over a longer time period. External evaluators could then be called upon at a critical point in time when impacts might be expected. Using both internal and external evaluators can also bring different perspectives to bear on the evaluation (26b).

Most of the identified evaluation approaches can be used for either forward tracking from a piece of research or backward tracking from a policy change or document, although the former (104) is more common than the latter (68). Hanney et al (65a) suggest that tracking forwards can identify a greater level of impact than tracking backwards. They suggest that this might be due in part to the reliance on self-reported data from lead investigators in forward tracking studies. However, there are some examples where evaluators have sought to track in both directions in order to construct a rich account of the relationship between research and policy. For example, the RAPID framework developed and applied by the Overseas Development Institute (71,82,87) draws upon episode studies of specific policy changes, case study analysis of specific research projects, and outcome mapping approaches. Information is collected using a range of methods including: literature review and documentary analysis; workshops with key staff, partners and actors; stakeholder interviews to triangulate workshop outputs; field visits; and a debriefing session to discuss initial findings.

One of the challenges of applying standard research methods is that they have not been developed for the particular purpose of evaluating the impact of research on policy. The same can be said for some of the data interpretation frameworks, including the various approaches to economic analysis, outcome mapping, benchmarking, simulations, episode studies and social analysis. A number of project management tools do include evaluation as a critical stage in the life cycle of a project. These include log frames (49,117), ROAMEF (Rationale, Objectives,

Appraisal in detail, Monitoring, Evaluation and Feedback) (37,56) and outcome mapping (52,87). The key strength of these approaches is that evaluation is embedded within the project management process, thus ensuring that it is not seen as a separate, stand alone activity, and is considered from the beginning of the project planning process. However, these different approaches do not provide detailed guidance on how to capture learning about processes and how to measure complex outcomes. Thus, despite their useful elements (see Table 1 in Appendix D), there are considerable challenges in applying these tools to the evaluation of research impact on policy.

Adopting and/or adapting a conceptual framework or model of the relationship between research and policy is, however, considered to be a valuable step in the development of a plan for studying the impact of research upon policy (100), and a number of models have been developed specifically for this purpose. These include the Health Economics Research Group (HERG) payback model, RAPID outcome assessment and the Research Impact Framework (RIF). For example, the HERG payback model explores five different impact domains: knowledge production, research targeting and capacity building, informing policy and product development, health (and health sector) benefits and broader economic benefits. There is much to be learnt from this approach, particularly as it has now been applied to a wide range of policy fields, such as health services and labour market research (65,144). The application of revised citation analysis techniques and a mixed method approach, combined with a clear conceptual framework (often used alongside a logic model to assess outputs and outcomes), makes the approach attractive in different countries and contexts. The RIF assesses policy impact around five dimensions: the level of policy-making, the type of policy, the nature of policy impact, policy networks and political capital. This framework generates brief one-page impact narratives that could be more attractive and applicable in applied research environments (84a). Brief details of each of these models can be found in Table 1 (Appendix D). A one page description of the RAPID and HERG approaches can be found in Appendix E and F.

4.3 Which methods are most effective?
Only 14 studies reflect directly on the effectiveness of different methods for evaluating the impact of research, and in only one of these was the effectiveness of

different approaches a central concern. Hanney et al (63) compared survey methods and case studies as methods for evaluating the impact of research investments. They conducted an empirical comparison of questionnaires and case studies, and found that questionnaires could provide a relatively robust approach to monitoring a broad research portfolio, but would benefit from a small number of accompanying case studies that would help to verify the information given and to test and improve the questionnaire for future use. This analysis was repeated with similar results as part of an evaluation of the impact of the NHS HTA programme (65a). The comparative study (63) emphasises the need to develop realistic approaches that fit with the scope of the evaluative task and the resources available.

In the remaining 12 papers, observations were generally limited to short paragraphs of text in the background or conclusions sections. These papers are consistent in their analysis, concluding that in order to gain meaningful insights into research impact, evaluators need to adopt in-depth, exploratory methods such as semi-structured interviews and case studies (143). Williams and Rank (142) believe case studies represent the best method for evaluating the impact of research and development, while Jones et al (77) conclude that there is no substitute for interviewing individuals, particularly within the policy arena. However, while agreeing that evaluators need to go beyond the use of surveys and quantitative indicators in appraising research impact, Anderson (6) argues that in-depth case studies can be prohibitively expensive.

Kilpatrick et al (79) consider the costs and benefits of a range of approaches to measuring research impact, noting that methods have been appropriated from evaluations in other fields or from more general social science research. They find that none of the 'borrowed' methods completely meets the theoretical needs of analysts assessing research impact, and conclude that while use of a combination of methods may overcome individual shortcomings, this is likely to be a more costly option. Other commentators also warn against the use of single methods, in particular, bibliometrics and econometrics (29) and argue that there is a need to find new (100) or 'hybrid' (143) methodologies

4.4 Which offer value for money?

The review found very little data on the value for money offered by the different approaches identified. At most, papers include short paragraphs reflecting on the methods used and their acceptability to participants. Typically (as with other research), the costs of conducting impact studies are not reported. The general theme emerging from reflective accounts seems to be that the most cost-effective and practicable approaches, such as surveys, are the least likely to yield insights into the process of research utilisation (particularly when conducted in isolation). There is a suggestion in one study (36) that telephone surveys might offer a cost-effective 'middle ground' between postal surveys and in-depth interviews: participants concluded that telephone interviews had offered an efficient use of their time as respondents, and had additional value as a method for identifying individuals for in-depth interview.

Two papers (38,63) in particular, conclude that there is unlikely to be sufficient time or resources to consider the impact of a whole portfolio of research. Evaluators will have to consider both issues of sampling (i.e. which research to focus on from within a programme) and the costs of potentially appropriate methods. For example, Hanney et al (63:189) conclude that while in-depth case studies generate valuable insights, they are 'inevitably fairly resource intensive and would not be appropriate for extensive surveying or regular monitoring of an R&D portfolio'.

Beacham et al (11) also suggest that the comprehensive approach supported by many commentators may well be too complex and expensive relative to the potential benefits, given the intricacy of the evaluative task. In this respect, it is perhaps worth keeping in mind some of the quicker, well established and more cost-effective methods, such as using an expert panel to provide an informed opinion about impact, developing a file to record personal testimonies, or developing proxy measures of impact. Davies and Dart (39) suggest consideration of the 'most significant change' technique, which focuses on collecting stories at the field level for discussion by a panel and collation in a report. Overall, the frequent recommendation to use a variety of evaluation methods is likely to add to the cost of impact studies. The suggestion that new methods need to be devised also has cost implications, given that methodological development work is time- and resource-intensive.

4.5 Are there any specific issues to consider when evaluating the impact of research on environmental policy?

The toolbox of methods used in evaluating the impact of research on policy seems to be relatively transferable to different policy domains. Hanney et al have conducted significant work in applying their model (HERG payback) to different policy domains. Based on their conclusions, two issues of transfer emerge. The first concerns the nature of the policy outcomes to be measured. Much environmental policy is less concerned with service delivery than with changes in attitudes and behaviour. Unlike health, there is no clear service sector or service delivery component to policy development. A second dimension concerns politically charged topics. In some policy areas, evaluators need to be very sensitive to the status of the topic in political terms. Some key debates may not feature in the paper trail of publicly available policy documents or may not be comfortable topics for policy makers to discuss. Finally, some areas are clearly more mature in policy terms and evaluators might therefore expect to identify more policy activity. Sustainability is a particularly challenging area of policy development, given its relative youth, its political profile and its remit to change 'hearts and minds' as a driver for change rather than focusing directly on providing services. These issues have more implications for the conceptual model of the policy process used to guide the evaluation, rather than for the methods used (although the choice of model is likely to have methodological implications).

Kivimaa et al (79a) suggest a further dimension of environmental policy that has implications for the evaluation of research impact. They argue that environmental policy typically involves a wide range of stakeholders and that the views of these stakeholders need to be heard in the evaluation process. They also highlight the inter- and trans-disciplinary nature of environmental research that needs to be taken into consideration in the design of research impact evaluation. Finally, a review of 33 organisations (70a) gave two specific examples of evaluations of environmental research impact on policy. Projects funded by the Finnish Environment Ministry are scored against a set of evaluation criteria by both the project leader and the Ministry supervisor. The Netherlands Environment Ministry has conducted two surveys (using external evaluators) of all policy makers who had commissioned research in a

particular year. The compulsory questionnaire included questions on how, and the extent to which, it had been used. The questionnaire was followed up with interviews.

5. What the literature doesn't tell us

The review found very few papers focusing on the costs and effectiveness of different approaches to evaluating the impact of research on policy. Only one study focused on comparing the effectiveness of two popular methods (case studies and survey methods). However, authors were more reflective about the advantages and disadvantages of different approaches and this information has been distilled in Table 1 (Appendix D).

The literature also identified fewer than expected empirical evaluations of the impact of research on policy. It is unclear whether this reflects a low level of activity, a failure to publish evaluations or the reporting of much of this activity in the grey (non-academic, report) literature that can be difficult to identify. The extensive internet search conducted as part of the review did provide access to grey literature so low numbers and lack of publication of evaluations may be more likely explanations. Those concrete examples of evaluations that were identified generated powerful insights relevant to this review and this literature is anticipated to grow in the future. At present, the majority of studies reflect on the relationship between research and policy rather than the impact of research on policy and, although this is an interesting and insightful literature, it was not the focus of this review.

Finally, while there is considerable reflection in the literature on the need to develop and test new methods for evaluating the impact of research on policy, or to modify methods used in other contexts, there are fewer empirical examples. To some extent, new methods are likely to be needed to reflect changing conceptual models of the relationship between knowledge and policy. For example, Gabbay et al (2003) have mapped the relationship between knowledge and practice as a mass of circles and boxes connected by complex webs of lines and arrows, which reflects the more realistic view of the non-linear processes of research influence and uptake, and represents a world apart from the neat circular and linear ROAMEF and log frames models. Innovations in technology, such as mapping software, make it possible to at least attempt to capture the complexity of policy making on paper. This is a field in

transition and there may well be applications of innovative methodological practice in development.

6. Conclusions

How do you (best) evaluate the impact of research on policy?

Although the literature is thin on empirical examples, it is full of reflective pieces. This is a field of practice currently experiencing rapid development and examples of innovative practice have recently been published (21,41,144). For example, new technological developments will now support more complex approaches to bibliometrics and network analysis. E-mail (79a) and the web (62) offer alternative vehicles for contacting and surveying participants. Eight key dimensions of best practice emerge from the literature analysis undertaken for this study. These are phrased as questions that need to be addressed in order to identify the most effective approach to evaluating the impact of research on policy. The literature stresses that the earlier these questions can be addressed in the research and development process, the better the evaluation will be (97,121). For example, good quality systems need to be in place in order to capture and track the use of research. Practical constraints in terms of time, budget and skills will inevitably guide any evaluation of research and development impacts on policy. Responses to the eight dimensions below can be used as a basis for developing an evaluation framework that is tailored to fit the organisational context, resources and requirements.

1. What conceptual understanding of the relationship between knowledge and policy is informing the evaluation? One way of thinking about this is to consider the current balance between different types of research use – enlightened, instrumental, interactive and political/tactical research (Weiss, 1979) – within the organisation and how this relates to where the organisation would like to be.

2. What are the outcomes of interest? For example, the HERG impact dimensions, to the door/through the door, positive and negative uses? To what extent is the focus on internal policy use or on a wider network of potential uses and users? For SCP these outcomes might be about changes in knowledge and attitudes rather than informing legislation or introducing new services.

3. What methods might be used to explore the outcomes of interest? For example:

   - achieving impact on policy documents (documentary analysis, citation analysis)

   - success of networks (network mapping)

   - being responsive to policy needs (interviews, documentary analysis)

   - involving policy makers in agenda setting (interviews, documentary analysis, workshops)

4. How does the evaluation address issues of attribution? Defra has expressed a particular interest in capturing the policy outcomes of research (as opposed to the outputs of the research process). However, there are implications for the degree of confidence it is possible to have in the results (see Figure 7). It is therefore critical that qualitative methods are used to understand the pathways and processes, and to describe perceived outcomes alongside attempts to develop meaningful outcome measures.

5. What is the direction of travel for the evaluation?

   - Backwards from policies/policy makers?

   - Forwards from research reports/research programmes/researchers?

   - Both directions?

6. Is this a mixed method approach, providing scope for triangulation? Are

   - The persuasive power of metrics, and

   - The explanatory power of qualitative methods

   used to best effect?

7. Will the methods selected capture the context and complexity of the research utilisation pathways, therefore helping to understand how (and whether) change has occurred? Will they capture changes in behaviours and relationships? There are valuable insights to be gained from the wider evaluation literature. For example, Theory Based Evaluation builds evaluation

from the logic behind a project or programme (Weiss, 1995) and Realistic Evaluation argues that a greater focus on understanding contexts and mechanisms will lead to more meaningful appreciation of outcomes. (Pawson and Tilley, 1997).

8. Does the timing of the evaluation offer sufficient time for change to occur, without compromising the likely recall capacities of respondents? Are there mechanisms that might be put in place to look at outcomes in the more immediate and longer terms? Is the evaluation conducted after a policy change has occurred (ex post) or is it predictive of potential use (ex ante)?

Demand for a more rigorous demonstration of the impact of research and development investments is growing. A National Audit report found:

> Evaluation of the quality of the research process is well established. However, there is a strong and developing emphasis on evaluation to encompass research relevance and value for money, as the link between research results and policy formulation increasingly becomes the focus of attention. (103: 2)

The evaluations identified in this review are consistent in their use of mixed method approaches, often using a combination of different qualitative methods such as in-depth interviews and documentary analysis. Approximately half of these are explicitly informed by a conceptual framework or model of the research-policy process. The descriptive and reflective papers argue that better evaluation in this field could be achieved. They argue for the development and application of conceptual frameworks (including, for example, HERG and RAPID but also encouraging the development of new models) and innovative methods for evaluation (advanced bibliometrics, developments in survey methods, network mapping and again encouraging the development of new approaches). Future evaluations might try to assess the impact of interventions designed to promote research use, such as knowledge brokers, networks, and broader linkage and exchange programmes. We are still a long way from providing a convincing answer to the question: does our investment in research and development make a difference?