How to evaluate the impact of environmental research on policy: LITERATURE REVIEW

Output from SKEP Call 2 - Science to Policy Process: Lot 1

Understanding the Impact of Environmental Research on Policy – Developing a Framework for Research Impact Evaluation and Guidelines for its Use

Policy Studies Institute

vito vision on technology

vrije Universiteit amsterdam

How to evaluate the impact of research on environmental policy: Literature Review

Version 1.1, May 2010

Output from SKEP Call 2 - Science to Policy Process: Lot 1
Understanding the Impact of Environmental Research on Policy – Developing a Framework for Research Impact Evaluation and Guidelines for its Use

Project Ref: 21205

This project was funded by the SKEP Network: a partnership of government ministries, environmental protection agencies and research councils responsible for funding environmental research. It is an independent network, supported by member subscriptions. Further information on the Network is available from www.skep-network.eu.

This Literature Review was written and edited by Ben Shaw and Sarah Bell of Policy Studies Institute, London with contributions from the project consortium consisting of staff from the Institute for Environmental Studies (IVM) (The Netherlands), *Technologie voor Ecologie en Economie (VITO)* (Belgium) and researchers working in a personal capacity: Dr Annette Boaz, King's College London, Dr John Holmes, University of Oxford and Louise Shaxson, independent science policy consultant.

If you have any comments or questions on this Literature Review please send them to:

Ben Shaw
Head of Environment Group
Policy Studies Institute
50 Hanson Street, London, W1W 6UP, United Kingdom
Email: b.shaw@psi.org.uk
Tel: +44 (0)20 7911 7544

How to evaluate the impact of research on environmental policy: Literature Review

Contents

List of Tables

List of Figures

1 Introduction

This Literature Review is an output from the SKEP Call 2 - Science to Policy Process project, 'An assessment of current approaches towards the evaluation of the uptake and impact of research projects and programmes by environmental policy-makers'. The aim of the project was to develop an approach for the evaluation of the implementation and uptake of environmental research and guidelines for its use by the SKEP Network.

The literature review aimed to synthesise and review existing knowledge and experience of the evaluation of research impact on policy and to use this as a building block for developing an applied model for environmental research evaluation impact on policy for use by SKEP partners.

An earlier version of the Literature Review was submitted in the Interim Report for this project. This revised version of the Review is a standalone document intended to provide more detailed material and references to support the main final outputs from the project: *How to evaluate the impact of research on environmental policy: a seven step guide* and the *Guidelines and Supporting Information* that accompany the Guide.

The review utilises and builds on the approach and findings from work that was conducted for the UK Department for Environment, Food and Rural Affairs (Defra) by two of the project partners[1]. This previous review addressed a similar research question to this project that is: how to best evaluate the impact of research on policy? The previous review was international in its scope, although the focus was on English language publications. The findings of the previous review, which was conducted in late 2007, have been updated, expanded and reworked to address the specific requirements of this project. Particular attention has been paid to publications released since 2007, to identifying non-English language publications and outputs from SKEP Network and similar organisations.

One of the findings to emerge at an early stage of the review is that there are few good examples or sources that consider the detail of the practical approaches that have been used to conduct research impact evaluations. We have therefore compiled a series of case studies to both better understand for ourselves and illustrate clearly to potential users the approaches that could be used to evaluate research impact. The case studies are contained in a separate standalone Case Studies report that accompanies this literature review.

The update of the previous Defra review has added value to this project through:

- Obtaining the most recent references in what is a fast evolving area of practice and theory,

[1] Boaz A, Fitzpatrick S, and Shaw B (2008) *Assessing the impact of research on policy: A review of the literature for a project on bridging research and policy through outcome evaluation.* Available at www.psi.org.uk/research/project.asp?project_id=182.

- Allowing the collation of more detail on research impact evaluation approaches and the issues associated with their use, and in particular the practical and conceptual/theoretical approaches used,
- Allowing us to seek material of specific relevance to this project and in which the previous review had found little material.

1.1 Review research questions

The overlap in objectives with this and the previous Defra review meant that similar research questions were used in this review. However, we expanded on these questions to address firstly specific interests in this project, and in particular an interest in the impact of research not just on policy but also uptake and implementation, and also areas in which little or less material was found in the previous review, such as the cost of different evaluation approaches. We sought material to answer the following seven questions:

1. How do people evaluate the impact of research programmes in different policy contexts and countries?
2. What are the strengths and weaknesses of different approaches to research impact evaluation?
3. Which methods are most effective?
4. Which offer value for money?
5. What are the issues and challenges for evaluating the impact of research on policy?
6. Are there any specific issues for evaluating the impact of research on environmental policy, regulatory uptake and implementation?
7. What conceptual and/or theoretical studies underpin research impact evaluation?

1.2 Review methods

The literature search and analysis was conducted drawing on principles from systematic reviewing, and the methods used are documented in the Appendices. The search process covered 11 literature databases (enabling the identification of journal literature, monographs, conference proceedings, papers and grey literature), internet searching of 42 websites, hand-searching of three important journals, citation follow-up and personal recommendations. Sources identified by these searches were selected for inclusion based on closer examination of how each source met the core review questions.

Whilst formal quality appraisal methods were not employed on the sources identified, the majority of sources had either been peer reviewed as part of formal journal peer-review processes, or represented the outcomes of large well-funded investigations using peer-review mechanisms such as panel review (e.g. by CAHS, CGIAR). Conclusions drawn from sources that did not fall into one of these categories were included only after careful consideration of

the methods employed in the research studies in question and with an awareness its potential limitations.

Data extraction sheets were completed to capture the relevant material (key findings, methodological considerations, basic reference information etc) from each selected source in a standard format. The data sheet template is given in the Appendices.

62 references were identified for this review building on the 156 references reviewed for the previous Defra review.

Efforts were also made to access non-English language literature, including searching European databases (such as FRANCIS), asking European project partners and SKEP members for recommendations of relevant literature, and re-running the initial Web of Knowledge searches to incorporate non-English language sources as appropriate given the language skills available across the project team (discussed further in Appendices)**.** This process proved particularly challenging, largely as a result of the limited project resources available, and the need to ask people to search on our behalf who were often less familiar with the literature of interest. The limited number of articles that were identified from these efforts had already been identified by the English searches which could perhaps be taken as an indication that the main methodological literature in this area tends to be in English.

Throughout this review, individual sources are referred to by a unique reference number with the full details being given in the References section. We have continued the numbering of references from the Defra review to avoid the potential confusion of numbers being duplicated. We have suffixed the numbers of all sources identified during this review with an 'S' (for SKEP), for example, 149S.[2]

The material identified in the review was first mapped to describe its basic attributes and character, and the material relevant to each of the project's research questions was then synthesised to address the seven questions of interest.

1.3 Structure of the review

This review is structured around the seven main review questions. Section 2 maps key characteristics of the literature sources synthesised in this review, e.g. aspects such as method used to identify sources, country of origin, themes addressed and research field.

Section 3 addresses the question, 'how do people evaluate the impact of research programmes in different policy contexts and countries'. This is the main section of the review and explores key questions to consider when planning an evaluation, the historical development of research

[2] The number of the first reference identified for this review is 147S, not 157S, as might be expected given the 156 references reviewed in the Defra review. This is because for cataloguing reasons some numbers in the Defra review were suffixed by letters, e.g. 26a, 26b which meant the last number used in the Defra review was 146.

impact evaluation, methods for collecting evaluation data, frameworks for structuring and interpreting data, and complementary evaluation tools and metrics.

Section 4 considers the strengths and weaknesses of different approaches to research impact evaluation. Section 5 provides a brief consideration of effectiveness, with Section 6 considering which offer value for money.

Section 7 considers the issues and challenges for evaluating the impact of research on policy, whilst Section 8 explores specific issues for evaluating the impact of research on *environmental* policy, regulatory uptake and implementation.

Section 9 highlights the range of conceptual and/or theoretical studies that underpin research impact evaluation, whilst Section 10 concludes with the key lessons to emerge from this review.

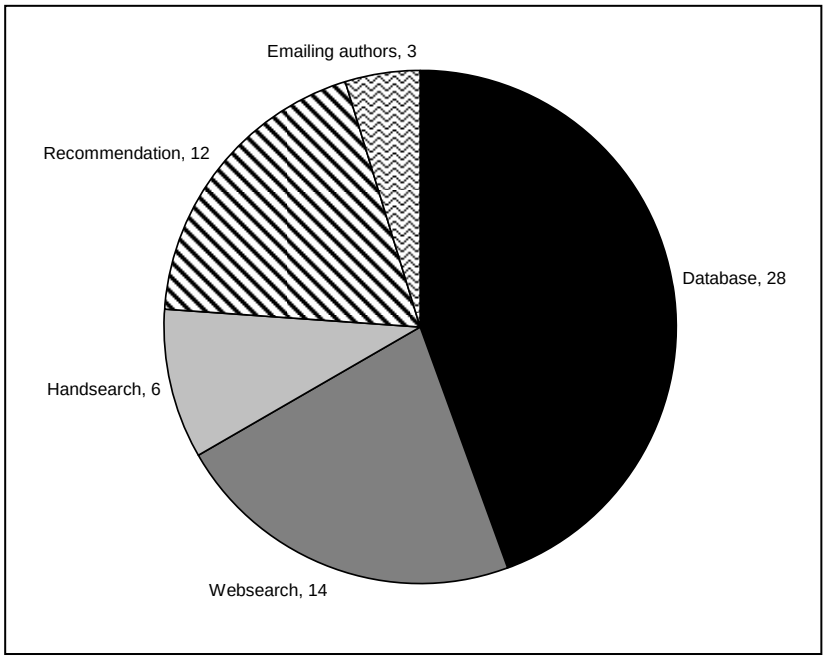
2 Mapping the literature

This and the Defra review have together considered 218 sources. This review identified 62 sources. The first stage of the analysis involved mapping and categorising the literature around a number of key variables: source identification method, country of origin, themes of interest to the review, research field, and the methodologies and frameworks used or discussed for conducting impact evaluations of research on policy. The results of this mapping are presented below in a combination of text and graphs. Further graphs have also been plotted:: to identify whether certain review questions were more commonly addressed in particular research fields than others; to look for changes in the extent to which different fields have focused on the review questions over time; and to establish whether there has been any increased activity in this area since he Defra review by comparing the dates of publication of the literature identified in this review with the dates of publications included in the Defra review.

2.1 Method of identification

The majority of papers were identified through database and web searches, but personal recommendations also made a significant contribution, particularly in terms of the quality of information provided (Figure 1). The 'emailing authors' category relates to references cited in other sources that were not available online or via database searches and therefore could only be obtained by contacting the papers' authors.

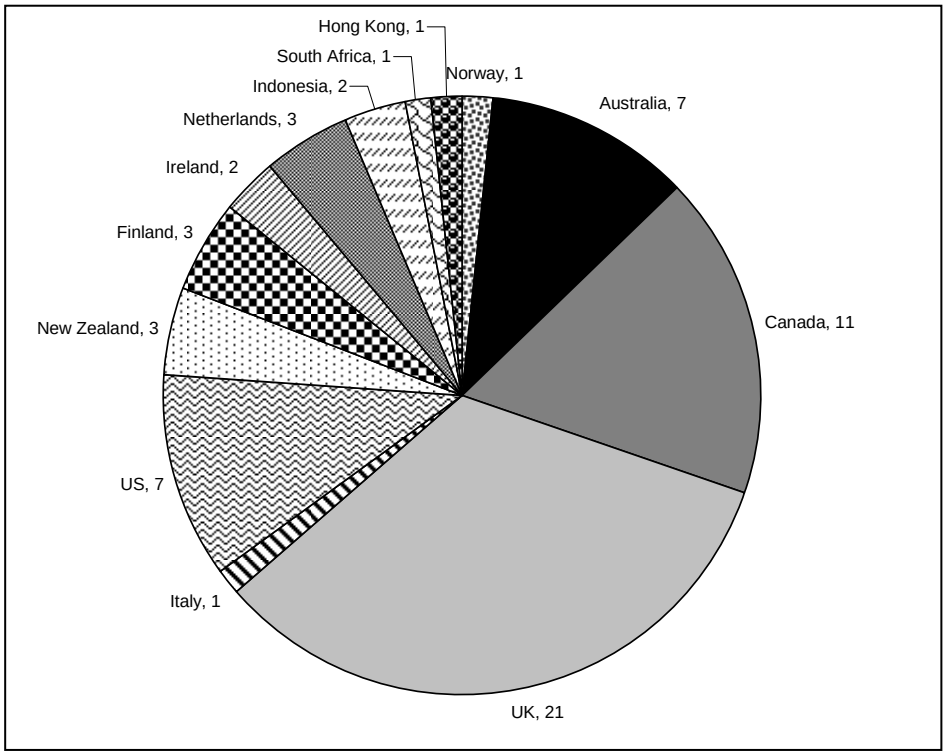
Figure 1: Review Sources - Method of identification



2.2 Country of origin

The majority of papers come from the UK and Canada, with smaller numbers from the US and Australia (Figure 2). Studies were also identified from a range of other countries, many European – including Italy, Finland, Ireland, the Netherlands and Norway – and some further afield, such as New Zealand, Hong Kong, South Africa and Indonesia. This spread of countries is likely to reflect the focus on English language publications. Efforts were made to access publications written in French, German and Dutch (according to the language skills at the disposal of the project team) but these are not included in this mapping of the literature.

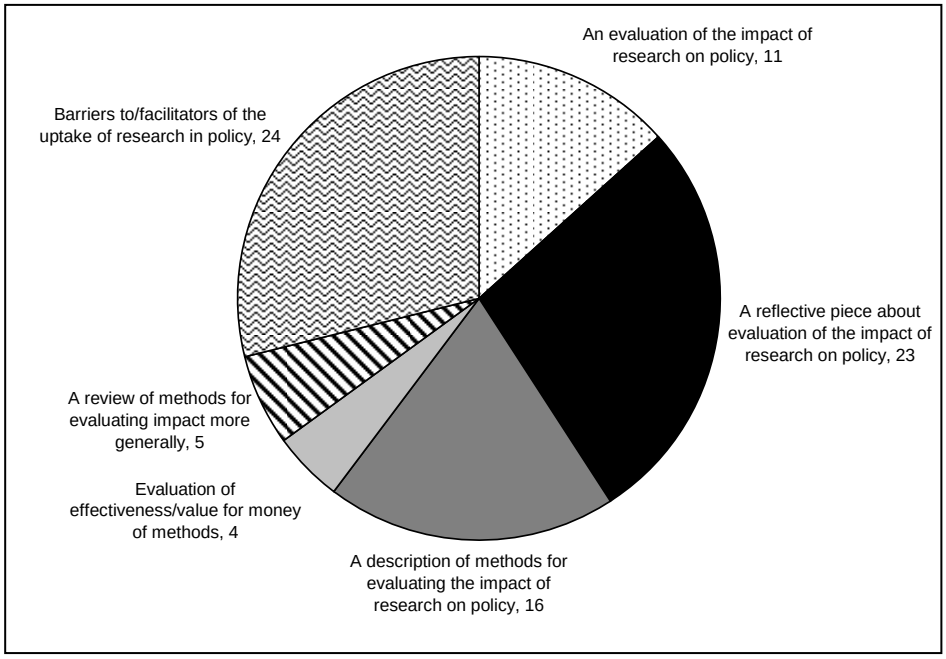
Figure 2: Review Sources - Country of origin



2.3 Themes addressed by sources

This review particularly focused on evaluations of the impact of research on policy. Of the 63 studies, 11 reported on empirical studies of the evaluation of the impact of research on policy and 16 described methods for evaluating the impact of research on policy (Figure 3). About a quarter of the studies were reflections on approaches for evaluating the impact of research on policy, and another quarter primarily discussed factors that commonly hinder or facilitate the uptake of research by policy-makers. Very few studies focused explicitly on the cost-effectiveness/value for money of different methods.

Figure 3: Review Sources - Themes addressed



2.4 Research field

The review identified a wide-ranging literature, reflecting a cross-sectoral interest in the issue of research impact on policy and evaluation of it (Figure 4). In particular there is a large health literature on the topic, as well as wider literature on evaluation and knowledge transfer. Particular effort was made to identify papers with topical relevance to the SKEP network, resulting in the identification of 18 environment and agriculture-related papers.

Figure 4: Review Sources - Research fields

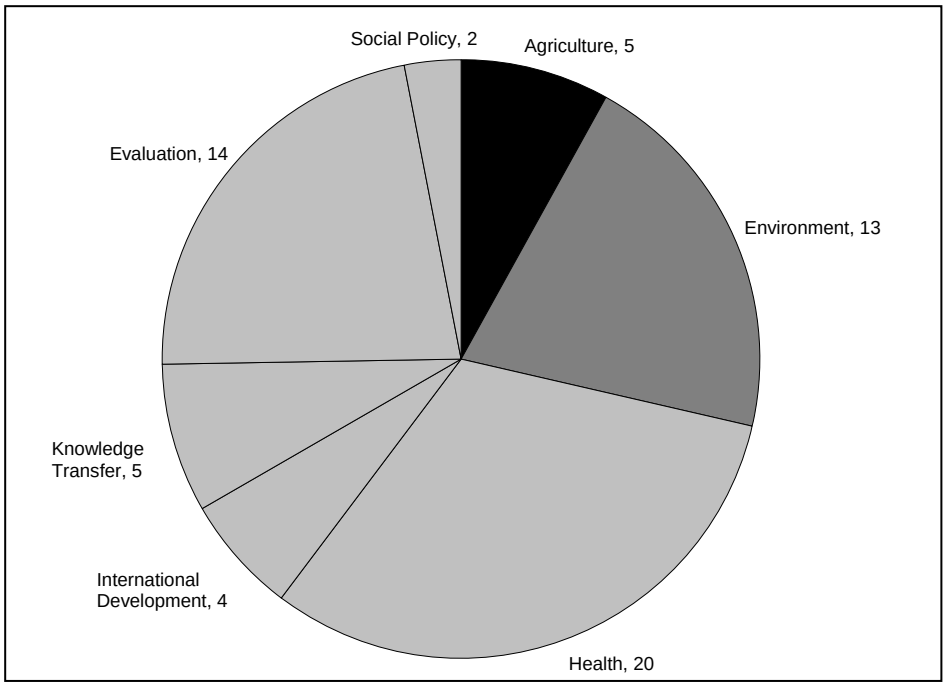


Figure 5 illustrates which of the review criteria of interest were most commonly addressed within different research fields. Empirical evaluations of the impact of research on policy occurred primarily in the fields of environment and health and, to some extent, agriculture, international development and social policy. Whilst the health literature seems to contain a number of reflective pieces, the environment literature more commonly just describes the methods in use. Perhaps this is a result of the longer-standing use of such evaluations in the health field compared with the environment field, as illustrated in Figure 6, such that the health literature has been able to move further beyond description to the reflection of their experiences with different methods.

Figure 5: Review Sources - Criteria of interest identified in each research field

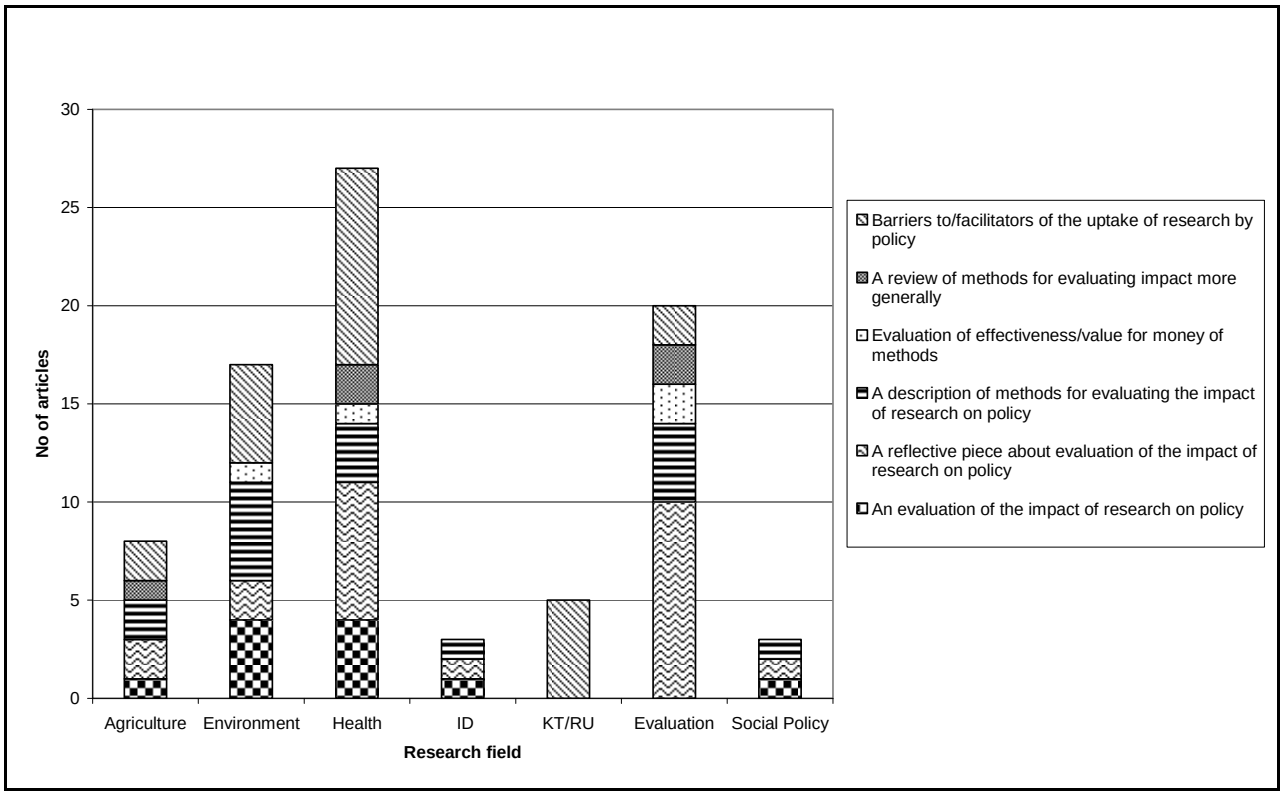
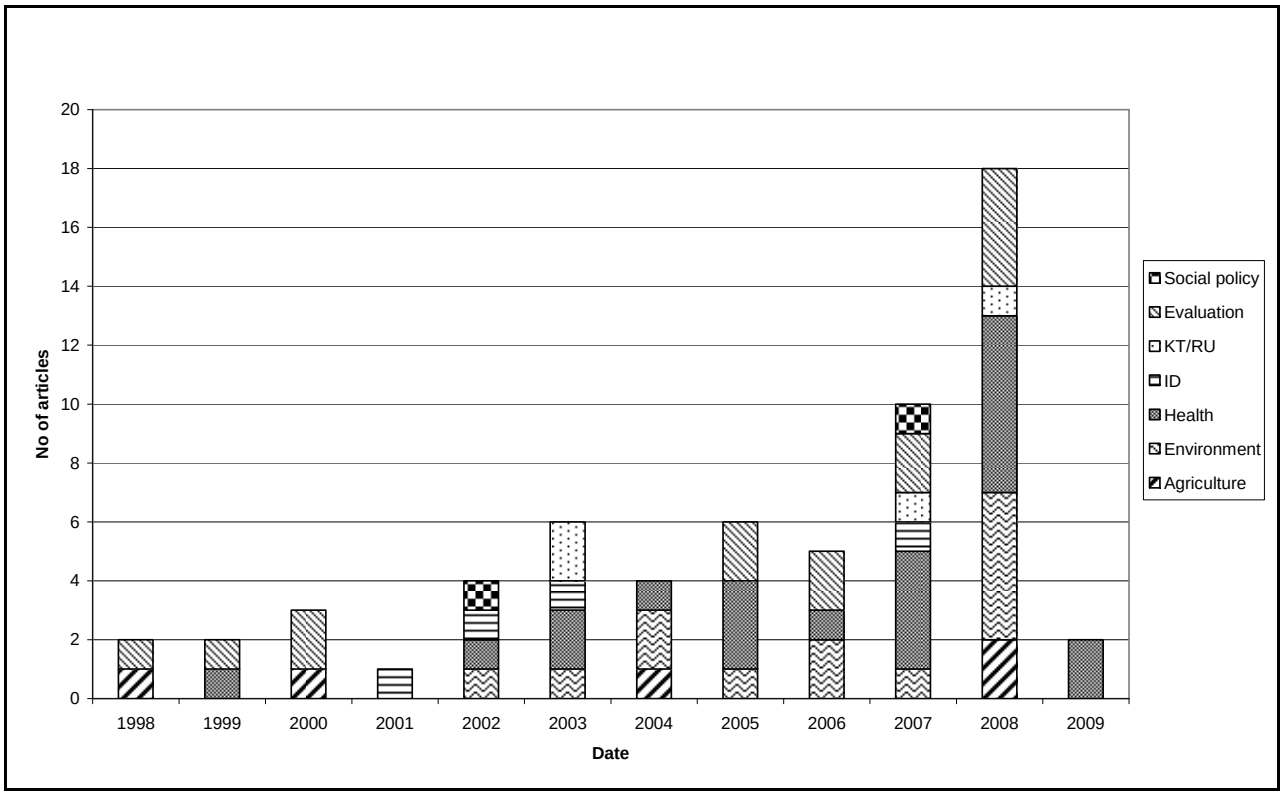


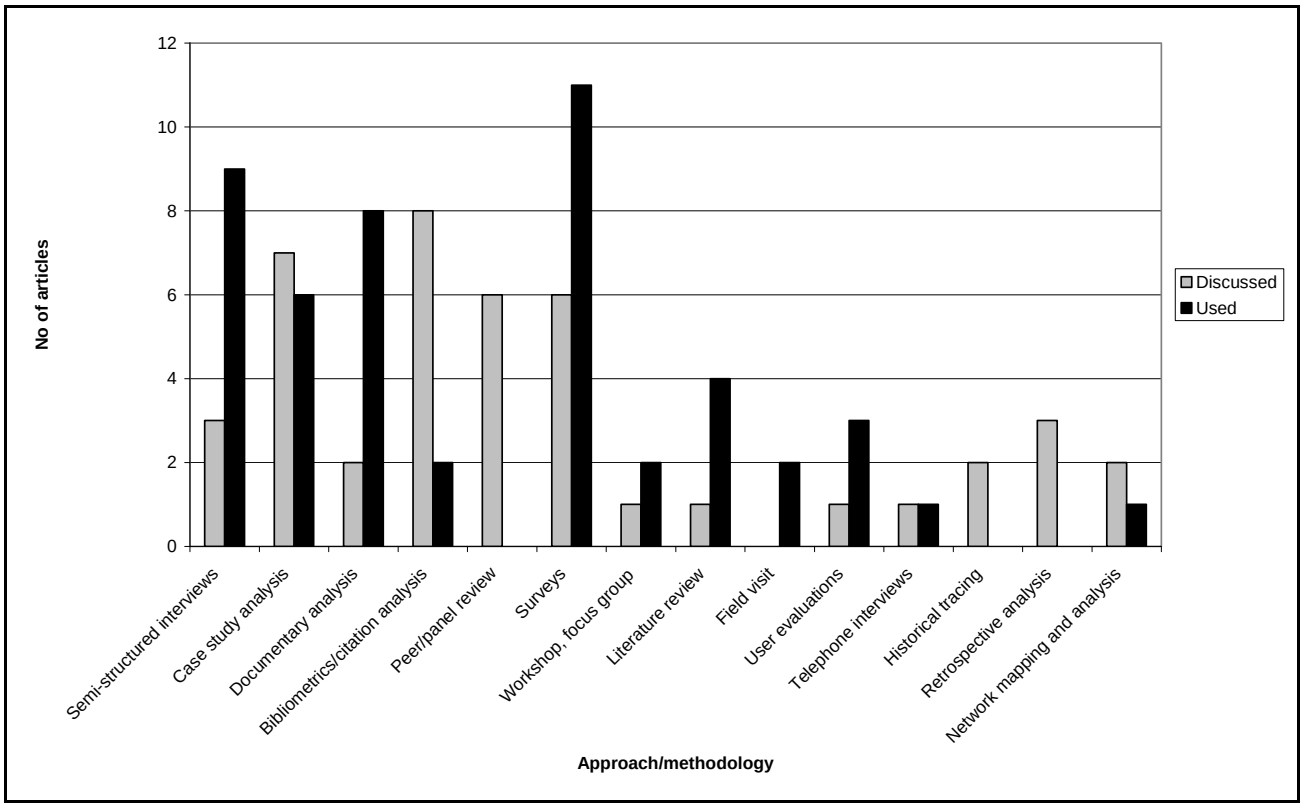
Figure 6: Review Sources - Frequency of articles published per year by research field
(N.B this graph does not include the earlier articles identified in the Defra review)



2.5 Data collection methods

A wide range of data collection methods have been applied to the study of research impact on policy (Figure 7). By far the most frequently occurring in the literature were surveys, semi-structured interviews, case study analysis, documentary analysis and bibliometric/citation analysis. When broken down into methods *used* and methods only *discussed*, it is apparent that the most frequently *used* methods are surveys, semi-structured interviews, case study analysis and documentary analysis, whilst the most frequently *discussed* methods are bibliometric/citation analysis, case study analysis, peer/panel reviews and surveys. Only case study analysis and telephone interviews were used about as much as they were discussed. There are many more references to, and uses of, qualitative methods than quantitative methods.

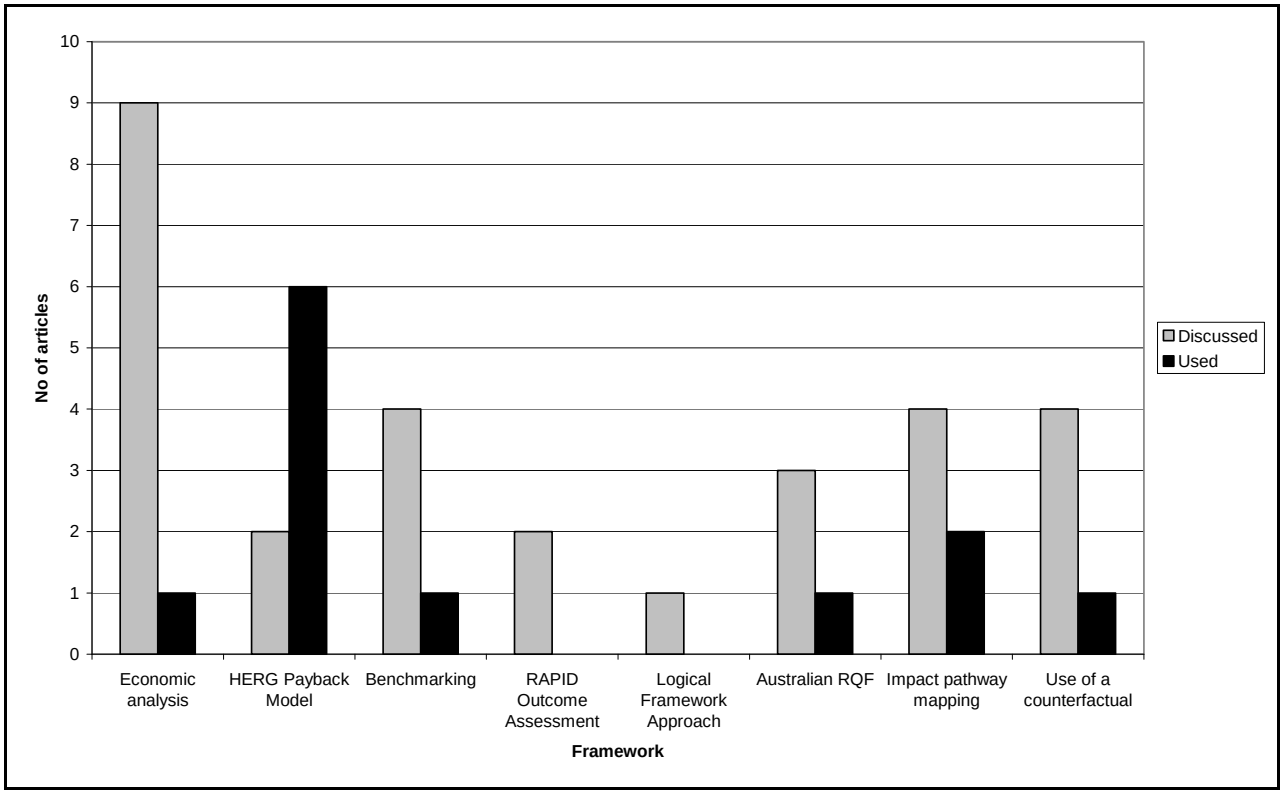
Figure 7: Review Sources - Frequency of use evaluation data collection methods



2.6 Frameworks and tools for structuring and interpreting evaluation data

A large variety of interpretation frameworks and tools are discussed in the literature but there actual use in impact evaluation (based on the availability of documented examples) is much less frequent (Figure 8). The most widely applied frameworks are the HERG Payback Model (particularly in health research) and Impact Mapping Pathways, whilst others such as economic analysis, benchmarking, use of a counterfactual and a research quality framework (RQF) approach have been used to some extent. Sometimes more than one framework or tool may be used in the same impact evaluation e.g. impact pathway mapping and the use of a counterfactual (CGIAR, 2008). (Brief descriptions of all the frameworks identified are given in Table 4 below in Section 3.)

Figure 8: Review Sources - Frequency of use of frameworks or tools for structuring and interpreting data

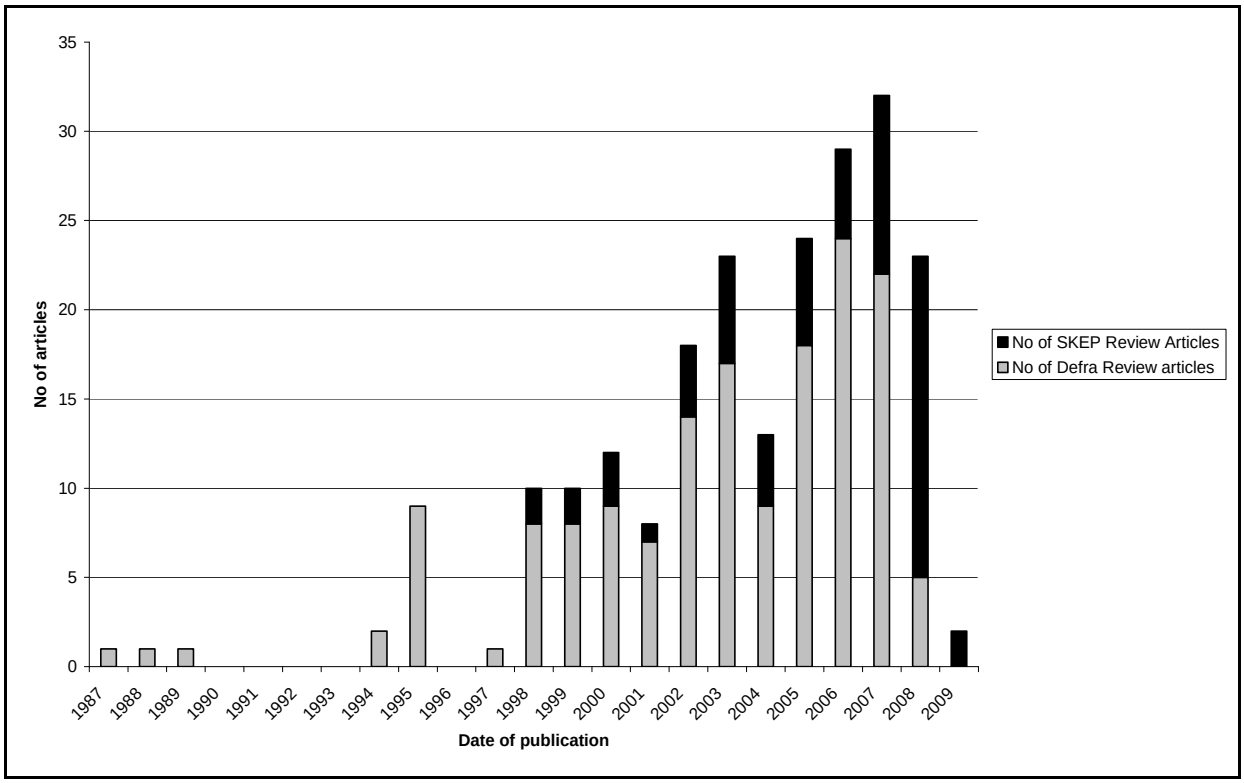


2.7 How much new literature did this review identify?

In conducting this literature review, we built on and expanded previous searches conducted as part of the Defra project 'Bridging research and policy through outcome evaluation'. We identified a significant number of new articles between 2007 and 2009, and a handful of articles from 1998 to 2006 not included in the previous review. This may be due to its wider scope and the decision to include articles which focus solely on factors that encourage or hinder research impact, which were not incorporated in the Defra review unless specifically linked to an empirical evaluation.

Figure 9 illustrates an increasing trend in the overall number of publications produced in this area until 2007, which decreased slightly in 2008 (as the literature searching for this project was completed in January 2009, this bar cannot be taken to represent the whole year).

Figure 9: Publication dates of the SKEP and Defra review sources



3 How do people evaluate the impact of research programmes in different policy contexts and countries?

The literature on evaluating the impact of research highlights a wide range of approaches and methods. This reflects the diversity of evaluation objectives and settings. It is regularly highlighted that evaluation approaches need to be tailored to the specific situation and evaluation objectives. This was noted in the previous review and is supported by the new material e.g. 151S, 152S, 155S and 193S. The type of research being evaluated may suggest or require different methods of evaluation. As Kivimaa et al. (79a) note:

> "Georghiou and Roessner (2000, 658) have argued that there is no universally applicable method for evaluation and that it is usually necessary to understand the setting of the evaluation and the discourse in which its results are located before the choice of approach can be fully appreciated."

This is representative of the sentiment in the broader literature and raises issues for this project in terms of proposing a prescriptive applied model of evaluation. Any model needs to balance sufficient detail to be of practical use, with flexibility to accommodate the range of settings and uses it might be applied in.

A large proportion of the new material we have reviewed focuses on what people could do to evaluate, guidance on how to do it, and/or the issues raised by evaluation (e.g. 149S). There is less material on what people have actually done to conduct evaluations that focus on the impact of research, although we found a larger proportion of case studies of evaluation than for the Defra study and also examples of methods not identified in it.

From considering the material synthesised in this review, there appear to be four distinct aspects to *how* people evaluate the impact of research. These are:

1. The process of planning for and developing an evaluation of research impact,
2. The methods that are used for gathering evaluation data,
3. The frameworks that are used to structure and interpret the data collected for evaluation.
4. The additional tools, metrics and indicators that can be drawn on and combined in order to augment existing, or develop new, evaluation frameworks.

In this section we discuss these aspects in turn and give examples to illustrate the main approaches used and summarise the strengths and weaknesses of different approaches.

3.1 Planning and developing a research impact evaluation

This project was intended to develop practical tools and guidelines for the development of research impact evaluation. As such we are interested not just in how evaluations might be conducted but also how they are planned and developed so that they are appropriate to the needs of a particular situation. The material we have reviewed highlights other sub-questions that evaluators, or commissioners of evaluation, need to answer before the question 'how do people evaluate the impact of research on policy?' can be answered. These are all necessary questions to address in the process of the development of an evaluation framework and include questions such as why, what, when and who do people or organisations evaluate. We initially consider what the literature we have reviewed has to say on these questions before moving on to discuss the actual methods used.

3.1.1 Why do people evaluate?

Most sources we have reviewed comment in some form on the importance of a clear rationale for conducting impact evaluation. These include justifications (from 79a, 149S, 151S, 161S, 181S, 193S, 196S and 208S amongst others) such as:

- To hold researchers, funding bodies and policy-makers to account,
- To direct and manage existing research and improve its value,
- To assist in setting future research priorities,
- To integrate research findings better with policy,
- To assess the ability and quality of researchers,
- To demonstrate the public value of research,
- To raise the profile of research,
- To understand the nature of relationship between research and policy, and
- To provide consistency and transparency in evaluation and of research and resource allocation.

The shift in motivation from a desire to evaluate research outputs and quality towards research outcomes and their impact is noted many times, both in terms of the desire to move beyond just considering research outputs, but also the availability of new or developing methodologies and technologies that allow this to be done (e.g. 149S, 181S, 193S).

3.1.2 What do people evaluate?

This question can be viewed from a number of perspectives, none of which are wrong or right; the approaches may simply be more or less relevant or useful to the context of the evaluation being considered or developed. These questions are also reflected in the conceptual approaches underpinning the evaluation which are explored in more detail in Section 9.

One approach to what is evaluated is the element of research that is to be evaluated, for example, (from 196S, although widely cited elsewhere in the literature), research inputs,

activities, outputs, immediate outcomes, intermediate outcomes, ultimate outcomes (examples of each of these are given in 196S, section 5.1). Consideration of these elements highlights a perhaps obvious, but easily overlooked point, that research impacts and evaluation of them is interested not just in research outputs and outcomes but also the inputs to research and all the activities that lead to the research outputs and outcomes and the links between them. However, this focus tends to suggest, implicitly or explicitly, a simple linear relationship between research inputs, outputs and outcomes, i.e. inputs lead sequentially and logically to outputs and then outcomes. However, the actual processes by which impacts occur are more complex and not necessarily linear as indicated, for example, in 196S in its comment on Bozeman's Public Value Mapping model which 'does not present a linear sequence of events. Instead, he utilizes the analogy of the "churn", meaning that scientific progress is not linear'.

There is also the type of research that is evaluated for, example, basic or applied. Basic research by virtue of its less predictable outcomes requires different evaluation approaches (e.g. 79a, 155S). 151S reviews eight research evaluation frameworks of which three were developed for the evaluation of applied research, four for basic research and one which is applied to both. Summary details of these frameworks are given in Table 2 later in the report.

Another perspective on what people evaluate is the *level* of research which is evaluated, for example, from an individual project, and up through research programmes, institutes, and national/international research funding bodies.

79a, (an output from a previous SKEP project) suggests four categories of programme evaluation (a higher level than the individual project evaluation) and these may cut across institutional and research project boundaries:

- **Programmatic** - considers the process, effectiveness or impacts of a research funding programme,
- **Thematic** - considers simultaneously a group of programmes' impacts on some specific outcome,
- **Knowledge systems** - focuses on evaluating the social configurations and the knowledge system around the research programme rather than merely scientific outputs, and
- **Research fields** - evaluating the development and knowledge on a particular research field.

Another way of looking at what is evaluated is in terms of the impacts of research that are considered, for example 197S reviews approaches to assessing the impact of policy-oriented research commissioned by Consultative Group on International Agricultural Research (CGIAR). It considers three types of impacts:

- Analysis of *diffusion* which refers to citations and other measures of the degree to which there is awareness and transmission of research findings among different audiences.

- Assessment of *influence* which refers to the degree to which the perceptions, conceptions and/or actions of policy makers have been changed (or confirmed) by research findings.
- *Impact* assessments analyse influence and attempt to quantify the benefits of changes in policy attributable to research.

It should be noted that 197S found few examples of assessments in the final category. Also that the situations in which CGIAR is interested in conducting evaluation and its objectives are very different to the European research content, although the methods it uses are relevant to this project.

One source also notes what has been evaluated less frequently:

> "…as the standard method progresses from inputs to outputs to outcomes in the literature, the results become less defined and the methods to measure them become more abstract….The links from immediate research outcomes to intermediate and ultimate outcomes are the least well studied but are of key interest to a small set of researchers focused on RIA [research impact assessment] methodologies, irrespective of subject matter." (196S, Section 6.0).

We return to this issue of attribution in Section 7.

3.1.3 When do people evaluate?

Evaluation can be done while the project is in progress or after it has been completed. Allowing a sufficient timescale for impacts to occur is a recurring point in the literature. There is a need to balance the allowance of sufficient time for impacts to have occurred against the likely decrease in the availability of people, documentation and other sources for involvement or consideration in an evaluation. We return to this issue of time-lags later in this section briefly and in more detail in Section 7.

The importance of considering evaluation at an early stage in the research development process and even before the research has started is also highlighted, for example, at the point at which the rational for research funding is set (208S). This is to build evaluation and resources for it into the research and research management processes, but also to give clarity on the research objectives for when an evaluation is, or might be done, at a later stage.

However, while clear research objectives may help evaluate the degree to which research has achieved its intended impacts, it is less helpful in the evaluation of unintended impacts of research. These can be a major part of the impact of any research, in particular basic or 'blue skies' research (e.g. 79a). 193S also highlights the problem of evaluating outcomes that were not defined at the outset of the work, in this case a lack of well defined societal goals for the Finnish Biodiversity Programme.

155S comments further on the differences between applied and basic research and suggests they both require different approaches and timing for outcome evaluation. It suggests that outcome evaluations could be conducted each year for applied research, as progress towards specific practical outcomes and milestones can be measured. It is much harder, however, to identify the practical outcomes of basic research in science and engineering, and annual evaluations would be unlikely to highlight its inherently unpredictable ultimate practical outcomes. Evaluation of the practical outcomes of basic research must therefore, it suggests, be retrospective and historical. In the meantime, in order to ensure that basic research programmes are generating the kinds of knowledge that have the potential to lead to valuable practical outcomes, annual evaluations could be conducted focusing on the quality of the research programmes, their contribution to world leadership in the relevant fields, and their relevance to agency goals and intended users.

3.1.4 Who is involved in the evaluation?

Kivimaa et al (79a) indicates that evaluations can be differentiated by who is doing the evaluation and suggests three types of evaluation: firstly, **expert evaluation** (whether consultants or peer review panels); secondly, **self-evaluation** (the researchers evaluate themselves); and thirdly, **internal evaluation**. Each approach has different strengths and weaknesses. For example, it is useful to undertake internal evaluation as it raises internal awareness of what does and doesn't work and also requires good levels of institutional co-operation and interaction. Alternatively, self evaluation is good for creating learning by researchers. However, both internal and self evaluation approaches are less good for understanding research outcomes as researchers and their funders/managers are not necessarily well placed to know the impacts that their research will have beyond their respective institutions.

A tension is also highlighted between peer review in it various forms, which is typically used to evaluate research quality, and professional evaluation performed by consultants. Professional evaluators are more likely to be able to effectively evaluate outcomes beyond the impacts created within the research discipline and can still consider research management processes and other impacts. However, both peer and professional review can deal with both.

155S also comments on this point and suggests that evaluations should be carried out by those with the knowledge and experience to understand its quality and, in the case of the applied research, its connection to public and agency goals. Applied research entails measuring some factors that basic research does not, such as ultimate usability, so the input of potential users is required, and that of experts who possess the ability to recognise its potential applicability to practical problems. This points to the need to consider who will supply inputs to the evaluation and broadly the literature indicates that researchers, research managers and funders, steering groups, research users and broader stakeholders (e.g. 193S) all have a role to play in outcome evaluation.

3.1.5 Key issues in planning an evaluation

These four questions: why, what, when, and who indicate some of the generic issues affecting the planning of a research impact evaluation indicated in the literature we have examined. They are presented here in a narrative form to familiarise readers with some of the issues in planning research impact evaluation. In the applied model of evaluation and guidelines developed by this project we aim to capture these points in a systematic framework and hope that this will enable those planning or commissioning evaluators to develop an approach suited to their needs and context. A number of sources examined however, suggest more formally the elements that need to be considered when planning a research impact evaluation.

For example, 79a suggests some key concepts of evaluation from the point of view of what needs to be evaluated as captured in Table 1.

Table 1. Key concepts and terminology associated with research impact evaluation.

Key concepts of evaluation	**General description**	**Description in the context of evaluating research funding programmes**
Target groups	The actors, i.e. decision making entities, such as companies, organizations and individuals whose actions the intervention is intended to influence.	Researchers in the organizations funded by the programme; other researchers, policy-makers and stakeholders that the dissemination of programme results is intended to reach.
Inputs	What is used to produce outputs. Resources, such as personnel and finance, but also matters coming from the target groups that the agencies take into account or respond to.	Money invested in research. Human resources used for planning, managing, advising and evaluating the programme. Management guidelines and research ideas and proposals from the different target groups.
Administration or process	The conversion of inputs, by an agency or a network, into outputs.	The conversion of money, personnel efforts, management guidelines and research ideas into the outputs defined below.
Outputs	What the administration produces and the target groups are provided with or expected to respond to.	Calls for research proposals, decisions for funded projects, conferences and seminars organised by the programme management and disseminating programme results, evaluation reports.
Outcomes (immediate)	The actions and the consequences of the actions taken by the target	Research results (publications, presentations, models & tools),

	groups due to responding to the outputs. Hardly any outcome is the results of some policy outputs alone, but instead is affected by a variety of factors.	knowledge of researchers and seminar participants, and new networks of people.
Outcomes (intermediate, final)		New policies, innovations (technologies, products, services, practices), new business models, patents, environmental improvements, public knowledge, etc.
Source: Table 1, Kivimaa et al (79a)		

Brutscher et al (151S) suggests that, in the context of health research, there are five key elements of research evaluation frameworks. These are:

1. **Evaluation objectives**, which flow from the four rationales of evaluation outlined: accountability; 'steering'; signalling; and advocacy;
2. **Outcome measures**, ranging from output measures, comprising the goods and services directly produced to impact measures, capturing the long-term changes research brings about;
3. **Levels of aggregation**, which may be low (in case of an individual researcher, for example), intermediate (in case of a faculty or research programme) or high (when a whole research discipline is evaluated);
4. **Timing**, which can be cross-sectional (if an evaluator is interested in the outcomes of one piece of research) or longitudinal (if the evaluator is interested in the outcomes from a research group over a certain period of time, for example, rather than a particular piece of research); and
5. **Evaluation methods**, comprising statistical data analyses, modelling methods (such as micro-econometric modelling) and qualitative and semi-quantitative methods (such as interviews and case studies).

These elements are drawn from a consideration of eight international health research impact evaluation frameworks (see Table 2 below for a summary of these frameworks). This work also considers whether there are interdependencies between these elements and therefore potential trade offs to be made between them when planning an evaluation. It concludes that there are likely to be trade offs, but that because of the small number of studies considered these need further investigation. It does indicate however, that that the choice of evaluation objective is 'immensely important' as:

> "The choice of objective(s) (when establishing a research evaluation framework) influences the choice of outcome measures, and the choice of outcome measures influences thinking about the right level of aggregation and timing. In addition, we propose that the level of aggregation influences the choice of methods".

These two sources have a different focus, 79a on the elements to be evaluated and 151S, on the issues an evaluation of these elements must take into account. These issues are highlighted in relation to planning an evaluation. Further points are noted below in relation to the use of the methods and issues these raise. The aim of this project is to bring these and other related points together in the applied model and guidelines.

Table 2. Eight health research impact evaluation frameworks

Framework -Country	Summary
Leiden University Medical Center (LUMC) – The Netherlands	This framework is an ex post evaluation framework which focuses on the 'societal impact' of research at the level of the research group. The framework can be seen as part of a broader movement in the Netherlands to correct for the 'serious imbalance' arising from the traditional focus of evaluation on scientific quality. The underlying assumption of the framework is that societal impact and scientific quality need not always go hand in hand. High quality research may not have immediate or long term impact and more practice-orientated research perhaps considered low quality by science may have immediate and important social benefits.
Measure of Research Impact and Achievement (MORIA) - Australia	This framework looks at outputs, outcomes and impacts of research across three domains: 'knowledge', 'health gain' and 'economic benefits'. It was developed as an analytic support instrument in the ex ante peer review process for grant applications. Similar to the LUMC framework, findings are translated into a standardized numerical score. This allows comparison and aggregation of findings across projects and within projects across different domains.
Program Assessment Rating Tool (PART) - USA	PART was introduced to improve government management and is used to assess the effectiveness of around 800 federal programmes. It takes the form of a 'diagnostic questionnaire'. An interesting element of PART is that to a large extent it evaluates programmes on the basis of performance goals. To do so, it adopts output, outcome and efficiency measures. Most weight is on outcome measures.
Vinnova (Swedish Governmental Agency for innovation systems) - Sweden	Vinnova is the Swedish Governmental Agency for innovation systems. Following its formation in 2001, there was an interest in understanding better what its initiatives were achieving, as well as in developing methods to estimate its long-term impacts. Since 2003, Vinnova has been conducting impact analyses of its work on a yearly basis. The framework consists of two main parts: an ongoing evaluation process and an impact analysis. There is some variation in how the framework is applied.
Payback (in use at the Canadian Institute of Health Research) - Canada	The Payback Model was developed at the Health Economic Research Group at Brunel University (HERG). It has been applied in a number of different contexts [as discussed elsewhere in this report]. The framework is an input-process-output-outcome framework. It comprises two components: a definition of evaluation criteria for outputs and outcomes of research and a logic model.
Department for Innovation, Universities and Skills (DIUS) - UK	This framework aims to 'assess the overall health of the science and innovation system, and how it delivers economic benefits'. It is the latest stage in a process of developing performance appraisal methods for the UK science and innovation system. The framework is used to model the delivery of economic impacts at the

	aggregate economy level through three stages and three influence factors.
European Union Framework Programme 7 (FP7) - EU	FP7 is meant as a key instrument contributing to the Lisbon, Gothenburg and Barcelona objectives – the system for evaluating the programme being a vector for tracking the results of research programmes and how they are contributing to the policy goals, and identifying what needs to be improved to achieving these goals more effectively. The responsibility for the evaluation of the Framework Programme rests with the evaluation unit in DG Research. It is supported by evaluation units in other DGs.
Congressionally Directed Medical Research Programs (CDMRP) - USA	These programmes are part of the US Army Medical Research and Material Command and CDMRP manages (some of the) biomedical research that US Congress assigns to the USAMRMC. The evaluation system consists of several elements the three main ones being: its grants management system, its product database and its (breast cancer) Concept Award Survey.
Source: adapted from Table 1, Brutscher et al (151S).	

3.2 The historical development of research impact evaluation

The shift from a focus on evaluating research quality to broader societal impacts of research is regularly noted (e.g. 193S). A number of sources highlight the historical development of methods for evaluating research impact mainly from the starting point of the use of bibliometrics as a measure, or proxy measure, of research impact on policy.

Lewison (179S) discusses the development of citation indexing over 50 years and highlights how this method could be further developed given advances in information technology, electronic databases and web-searching. Citation analysis can now be used for research outputs cited beyond the formal academic literature, for example in national or international standards, policy documents and newspapers, thereby giving an indication of the possible influence or impact of research beyond the academic community and literature. 204S highlights one prospective means of measuring the impact of research information by developing an ‘online feedback loop’ with users of health research. This suggestion was published in 2001 and, of course, the possibilities of internet-based approaches have grown enormously since them. 208S highlights an example of an evaluation project combining the use of Google searches to identify project information in the public domain with tailored surveys of researchers, end of award reports and interviews.

155S suggests six methods for evaluating research were in use in 1998: bibliometric analysis, economic rate of return, peer review, case study, retrospective analysis, and benchmarking and reviews the strengths and weaknesses of these. The number of frameworks and methods found by our work suggests a rapid development of research evaluation approaches in the last ten years (or an incomplete consideration of available methods in 155S). 160S describes three waves of attempts to capture the public value of scientific research to date: a) quantitative

efforts (the emergence of sociometric aspirations); b) combining qualitative and quantitative data to measure outputs that contribute to outcomes (rather than outcomes themselves); c) a more qualitative approach, combining measures of research quality and impact with peer and user evaluation, to produce a logic model of the research process.

151S indicates the period over which the eight evaluation frameworks it reviews have been in regular use and notes that three-quarters of the methods it considers have been developed in the five years preceding the paper's publication (2008) and that this is indicative of a growing momentum around research evaluation in the last decade.

It is clear from these examples and the previous review that the area of research impact evaluation is evolving fast.

3.3 Methods for collecting evaluation data

This sub-section considers the methods used to collect evaluation data. The following sub-section considered the frameworks and tools used to structure and analyse the collected data. There is some overlap between the methods in these categories and the evaluation tools as the boundary between data collection methods and analytical tools is blurred. In practical terms this points to the need to consider both the analytical uses of the tools and the data requirements to conduct them when planning an evaluation.

This review, together with the previous Defra review, has identified 13 distinct methods for gathering data for research impact evaluation. Table 3 below briefly describes the basis of each of the methods, lists the references that cite them, and indicates the strengths and weaknesses of each method. The 13 methods can be grouped into five broad categories:

1. **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 of research and the policy context in which utilisation is, or is not, taking place. However, they are often time- and resource-intensive, and it can be difficult to generalise their findings.

2. **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 relatively inexpensive. 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 output, rather than its quality, and research *outputs* rather than *outcomes*.

3. **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. However, panels are open to accusations of a lack of objectivity and are heavily reliant on the quality of their membership.

4. **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. It is, however, a relatively quick and low cost option for evaluation. Examples of the value of workshops are provided by the UK Environment Agency approach (see Case Study 7 in accompanying Case Studies report).

5. **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.

This wide range of data collection methods obviously allow very different approaches to be used to conduct evaluation and many of these approaches are used alongside one another in a single evaluation. An example of a mixed method approach is provided by the Foundation for Research, Science and Technology (FRST) in New Zealand and its Natural Ecosystem Research Evaluation (see 163S and Case Study 6 in accompanying Case Studies report). Their approach aimed to evaluate the outcomes of the natural ecosystem research funded by FRST, based on the views of people involved in the management and/or protection of New Zealand ecosystems. Views were sought on: the benefits of natural ecosystem research, relationships with researchers, and factors affecting the implementation of research results.

The FRST evaluation approach consisted of two parts; case studies of ecosystems research programmes, mainly focusing on a researcher perspective, and a user survey with two components: a web-based user survey to give breadth of views, and interviews to give depth. No conceptual justification for the approach taken by FRST is made explicit. However, in practical terms the approach appears to have resulted in a clear indication of some of the positive impacts of the FRST research based on the views of a sample of users. It also identified issues for evaluation – the attribution issue (which is very commonly cited) and the fact that a research project is unlikely to be relevant to a wide community of users.

Similarly, 193S discusses the evaluation of the Finnish Biodiversity Programme, FIBRE, a six-year €20 million programme funding 105 research projects. Data for this evaluation were collected from several sources to allow management, programme and project level impact evaluation. Management factors such as administrative structures, measures of organisational involvement and interaction were examined by interviewing members of the steering committee and the co-ordinating office and by analysing the minutes of the committee meetings. The programme level parameters such as goals, research benefits, networks and dissemination and impacts were examined by interviewing steering committee members, end users and funded researchers. An email questionnaire was sent to the doctoral students involved in FIBRE projects.

A separate more detailed report[3] on the FIBRE programme evaluation highlights the benefit of a mid-term evaluation as a mechanism that can direct the subsequent stages of a research programme and, in this case, provide an increased focus on dissemination activities. In this case it was conducted by an evaluation panel which met with all 36 consortia / projects within the FIBRE programme and held discussions with stakeholders on the steering committee. It was considered an expensive but important process in shaping the subsequent activities and effectiveness of the programme.

Table 3 below briefly describes the basis of each of the data collection methods, lists the references that cite them, and indicates the strengths and weaknesses of each method.

[3] Merja Otronen and Juhani Tirkkonen, The National Impact of the Finnish Biodiversity Research Programme Fibre. Available from http://www.Aka.Fi/Fibre/Final/Impact.Pdf. 193S is based on this longer report but a data extraction sheet has not been completed for it.

Table 3. Strengths and weaknesses of different methods for gathering evaluation data

Method **Number of refs** **Reference number**[4]	**Description**	**Strengths**	**Weaknesses**
Semi-structured interviews: **Total 12** 02, 06, 11, **13**, **14, 16, 20**, **21**, **24,** 25, 28, **30, 31**, **36**, **45**, 46, **47**, **50**, **53**, **54**, 55, 56, 57, **58**, 59, 61, 64, 65, **68**, **70a, 71**, 73, **74**, **77**, **84a, 85**, 86, **87**, **88**, 92, 93, **96a, 101, 109**, **111**, **115**, **117, 121, 123**, **124**, **127, 133, 136a, 140**, **143**, **144**, 146, 148S, **151S**, **153S, 162S, 163S, 169S**, **197S, 176S, 193S,** 196S, 197S, **199S**	A flexible interview around a framework of themes, with pre-identified key actors.	• Flexible structure enables interviewees to respond in own terms, and interviewer to respond as part of a two-way conversation • Framework ensures comparability of interviews • Seen as most appropriate when unravelling diverse layers and subtle nuances • Useful when there is a limited population of individuals with significant enough breadth and depth of knowledge to speak informatively about what happened and why (key informant interviews)	• Issues of bias and attribution • Training for interviewer is necessary (to avoid pre-empting answers or not probing enough) • Time-intensive (collection and analysis) • Interviewees may themselves be unaware of indirect influences of research • Time recall • Difficulties in locating appropriate key informants/research users • Often poor record-keeping by researchers of activities undertaken that may have encouraged research uptake (ephemeral connections), and by research users of specifically when/how they used the research
Case studies: **Total 68** **03, 06**, 10, 11, **12, 14**, 15, **16,** 17, 18, **21**, **23**, 25, **26, 26a, 26b, 26c, 26d**[5]**,** 28, 29, **30,** 38, **40**, **45**, **47**, **51**, **52, 53**, 65, **65a**, **67**, 72, 73, **79a, 85**, 86, **87**, 92, 93, **96a**, **101**, **104**, **105, 107**, **109**, **114**, **121,** 125, **126,** 132, **133**, **136a**, 137, 142, **143**, **144**, 146, 151S**,** 152S, **153S,** 155S, **156S, 161S, 163S,** 170S,	An empirical approach that explores a project/programme in-depth, describing and explaining how and why developments of interest have occurred.	• Can be descriptive and explanatory, and rich in detail • Can demonstrate pathways from research to impact • Potential to combine sources and methods (triangulation) • Explores context • Particularly useful in situations and contexts where the understanding of research impacts is poor, and there is only some knowledge on the causalities involved	• Difficulty in selecting cases • Issues of bias and attribution • Often over-emphasises importance of research – 'supply perspective' • Time-intensive to ensure rigour • Difficult to apply a common framework across case studies. Must follow a standardised protocol in order to be comparable. • Often poor generalisability of findings • Highly dependent on the skill of the researcher

[4] A reference number in bold typeface signifies studies that use the method, non bold signifies studies that only discuss the method.
[5] There are multiple references to one particular study conducted by IDRC (26, 26a, 26b, 26c, 26d).

185S, 196S, **207S,** 208S			
Documentary review: **Total 55** **02**, **03**, 09, 11, **14, 16, 20**, **21**, **30,** 41, **44**, **45**, 46, **47**, **50**, **52**, **53**, **54**, 57, **58**, 64, 65, **65a, 68**, **71**, **74**, **79a, 84a, 85**, 86, **87**, **96a, 100**, **104**, **109**, **111**, 112, **115, 117**, **123**, **127**, **133**, **140**, **143**, **144,** 148S, **153S, 157S, 169S, 187S, 192S, 193S,** 196S, **197S, 202S**	Exploration and interpretation of existing documents. Can elicit quantitative or qualitative findings. Often used in conjunction with other methods.	• Can be applied to a wide range of sources (including policy statements, technical reports, minutes, speeches) • Provides contextual understanding • Potentially cost-effective	• Relies upon the quality of existing records and access to these • No single methodology for analysis
Bibliometrics/citation analysis: **Total 47** 09, 11, 15, 18, 21, 23, 28, 29, 30, 34, **36**, 41, 52, 56, 61, 64, 65, 72, 73, 76, 78, 79, 79a, 80, 81, 82, 97, **96a,** 98, **104**, 106, 127, 131, **133, 134**, 137, **143, 144**, 151S, 152S, 155S, **169S**, 170S, 179S, 180S, 196S, **202S**	Method for quantifying the impact of research by counting the number of outputs and citations, and analysing citation data.	• Suitable for repeated analyses and comparisons • Measures original research, not the programmes resulting from it • Identifies research deemed to be important by subsequent research • Traditionally considers academic citations, but can be extended to grey literature, policy documents and media • Data easily available • Relatively inexpensive • Provides rapid results • Correlates with other evaluation methods	• Measures outputs not outcomes; citations by researchers only represent the first of many steps by which research is put to use by society. • Not comparable between disciplines • Quantity of output may not reflect quality; risks distorting importance of select publications • Limited role where publication is not the goal of research • Many uncertainties, and thus can only be used as a partial indicator • Time lag between publication and citation • Can be artificially influenced

Peer/panel review: **<u>Total 38</u>** 09, 11, 15, 18, 28, 29, 37, 38, 41, 48, **50**, 55, 56, 60, 73, 78, 79, 80, 81, 82, 83, 92, 93, 97, 102, **105**, **109**, 120, 135, 137, 142, **143**, 151S, 152S, 155S, 170S, 196S, 208S	Widely used advisory process of expert scrutiny of projects and programmes. Traditionally used to enhance or maintain quality of science, on the basis that experts in a particular field can reach a consensus.	• 'Experts' confer status, credibility, and acceptability on findings • Can offer range of constructive feedback to guide process • Can be conducted at any time during the process of impact analysis • Relatively cost-effective • Flexible • Well understood • Useful for evaluating fundamental, long-term projects whose ultimate outcomes are unpredictable and not easily quantified • Can determine both quality and relevance of research	• Reflects upon rather than measures impact • Time-consuming method, particularly for the experts involved • Issues of objectivity and variability (including the quality of experts) • Impractical to evaluate a broad area (number of peers involved) • Criticised as eliciting acceptability rather than validity of findings • Qualitative findings only • Requires comprehensive information for reviewers
Surveys: **<u>Total 47</u>** **06**, 09, 11, 15, **16,** 18, **21**, **23**, 28, **31**, 55, 56, 64, **65a,** 73, 74, **75**, 78, 79a, 86, 92, 93, **96a, 101, 111, 115**, **133**, **134**, 137, 142, **144, 150S,** 151S, **152S, 153S,** 157S, **163S,** 170S, **173S, 174S, 176S, 183S, 187S, 192S,** 196S, **202S,** 204S	A pre-formatted series of questions asked of multiple actors, generating both quantitative and qualitative data.	• Can identify outputs/outcomes associated with particular research • Cost-effective means of obtaining data from a range of actors • Allows wider involvement range of stakeholders than possible with interviewing • Can identify greater breadth of impact • Can identify aspects to focus on in interviews • Useful for the triangulation of findings from other methods	• Relies upon access to respondents • Reflects the bias of those surveyed and those who respond • Unresponsive to unforeseen issues • May require follow-up interviews to fully understand the results • A questionnaire approach can overlook nuances and subtleties surrounding influence • Can suffer from low response rates
Workshop, focus group: **<u>Total 23</u>** 03, 18, **19, 24,** 25, 28, 29, 42, 43, 49, **53**, **58**, **71**, **79a, 87**, **96a, 115**, **121, 127,** 135, **187S, 195S,** 204S	An organised discussion with a group of individuals. The groups can involve a range of different stakeholders.	• Can be conducted at any time during the process of impact analysis • Less expensive than surveys • Exploratory and in-depth insights possible • Can reach a consensus that individual responses may not	• Risk of sample bias • Selective memory of participants • Unsuitable for competitive or sensitive topics

Literature review: **Total 23** 12, **14,** 16, 25, 29, **30, 31**, **45,** 65, 74, **79a**, **87**, **104, 111**, **115**, **127, 143, 144, 148S, 177S,** 184S, **196S, 198S**	Synthesis of existing research relevant to the study. In the context of impact evaluation, usually used with other methods.	• Useful initial research to define the scope of an impact study/ look at evaluation methods • Cost-effective	• Depends upon ability to identify and access existing research
Field visit: **Total 15** 09, **16,** 42, 43, **45, 47**, 48, **50**, **71**, **87**, **121**, 127, 135, **162S, 176S**	Primary research method where the research team visits in person the site of activity. Often used in the international development field to evaluate the broader impacts of a research programme. Can include observing meetings.	• Can ensure information is up-to-date • Direct observation of activities	• Time-intensive and costly (requires planning and ex-post evaluation if it is to be beneficial) • Distorts on the ground activity
User evaluations: **Total 12** 40, 72, 78, 79, 96, 97, 120, **140,** 148S, **163S, 187S, 202S**	Participatory method for assessing stakeholder satisfaction (either users of research or producers of research). May involve interviews, e-mail or telephone survey.	• Looks at both research and research utilisation from a stakeholder perspective • In-depth understanding of utilisation processes	• Risk that stakeholders have a vested interest in expressing satisfaction • Can be both time-intensive and costly
Telephone interviews: **Total 10** **06**, **36,** 74, **79a, 85**, **88, 100, 104, 150S,** 182S	Usually semi-structured interview, often used as preliminary means of identifying key stakeholders.	• Can elicit open-ended information early on in research process • Enables greater sampling dispersion • Cost- and time-effective	• Less able to develop rapport with interviewee (restricts line of questioning) • Interview length is limited • Cannot use visual prompts/sources
Patents/new technologies: **Total 6** 11, 18, 34, 41, 76, 82	Where research may have patentable outcomes, this approach gathers data about the number and nature of patents.	• Useful to identify linkages between research and specific outcomes	• Difficult to compare between disciplines • Time lag after publication • Most commonly used to evaluate the wider impact of research on industry
Impact Log: **Total 3** 11, 38, 72	A means of logging real-time direct impacts and uptake of research (i.e. from informal feedback, field observations).	• Records impacts over time • Can be used to construct actor matrices	• Cannot be conducted retrospectively • Much of the information is anecdotal and subjective • Does not measure indirect impact

3.4 Frameworks and tools for structuring and interpreting evaluation data

Multiple data collection methods are normally used within a broader framework for structuring and analysing the data gathered. As indicated by 151S, and many other references, multiple data collection methods tend to be used in a single evaluation framework and these mix quantitative and qualitative approaches. Of the eight international examples of evaluation frameworks considered in 151S, seven out of eight used at least three different data collection methods, and two used six methods. Only one of the eight frameworks relied on a single data collection method. This use of mixed methods approaches is seen across the literature and practice more broadly.

This section initially summarises some of the key frameworks that have been used or proposed for evaluating the impact of research, and will then introduce additional evaluation tools that have been used within or alongside these frameworks in order to tailor them more specifically to the needs of the organisation commissioning the impact evaluation. This section is intended to illustrate the wide range of approaches in use. The nine case studies given in the accompanying Case Studies report are also relevant in terms of illustrating the detail of some of the approaches used.

An important aspect of some of these frameworks highlighted by a number of authors is the importance of using conceptual models of the relationship between research and impacts e.g. 149S, 170S, and in particular 202S. Whilst some comments are made on conceptual approaches in this section, a more detailed discussion of the conceptual and theoretical basis underpinning these evaluation approaches is provided in Section 9.

This and the Defra review have together considered 218 sources and from this we have identified 24 different evaluation frameworks, tools and metrics that are used to structure and interpret research impact evaluation data. These are listed in Table 4 and the basis of each approach summarised briefly.

The approaches vary widely in terms of their input data requirements, outputs, scope of analysis, purpose, timing, expertise requirements, rigour and resource implications. Table 4 indicates the strengths and weaknesses of each these approaches which we have identified from the literature and the references that cite and comment on the approach. The data extraction sheets and the references themselves contain further detail on these approaches.

Table 4. Strengths and weaknesses of different frameworks and tools for structuring and interpreting evaluation data.

Framework **Number of refs** **Reference numbers**[6] **Summary Description**	**Strengths**	**Weaknesses**
Research impact evaluation frameworks; rounded evaluation approaches that are adaptable to a range of situations, based on mixed method approaches, underpinned by a conceptual model of the research-policy relationship. Includes:		
HERG Payback Model: **Total 20:** 11, **23**, 29, 38, **65a,** 84, 97, **104**, **114**, **133**, **143, 144**, **152S,** 160S, **169S, 173S, 185S, 191S,** 196S, **207S** A comprehensive framework for recording the wider impacts of research across five domains: services; policy; practice; research; and capacity-building. Incorporates qualitative and quantitative data collection.	• Common framework enables comparative analysis of multiple projects • Explores and explains impact • Mixed-method approach to counteract bias • Comprehensive and systematic • Potential to score observed payback/benefits (although methodologies for this need further refining) • Applied and tested in a wide number of impact studies	• Costly and time-intensive (breadth of information) • Assumes all outputs have equal impact • Multi-dimensional categories risk double-counting and can make interviews repetitive • Does not fully explain/account for complex research-policy interface and non-linear impact of research on policy. • Developed and primarily applied in health care research
Research Impact Framework (RIF): **Total 2:** 84, **84a** An easy-to-use framework that includes descriptive categories to prompt researchers to systematically explore and describe the impact of their work.	• Standardised structure that can be used to compare projects and programmes • Maps potential impacts (positive and negative) which can be used to clarify dimensions of use • Simple, quick and low cost approach	• Descriptive categories – analysis remains subjective • Lack of triangulation • Developed and applied in health care research
Australian Research Quality Framework (RQF)/Irish EPA approach: **Total 4:** 149S, 159S, 169S, **178S** Two similar indicator based approaches that uses expert	• Assess the quality and impact of research • Can alter or add to criteria used as programme requirements evolve • Includes a context statement to account for impacts not covered by the criteria i.e. adds flexibility • Allows comparisons across research projects /programmes by including generic criteria, whilst	• Rating of impacts by expert panels can be subjective • Rating methodologies may need to be refined. For example, temptation to rate a project with a 'very good' impact on the basis of potential rather than actual impact • The indicators could become the sole focus of attention, whereas they are meant to act as a guide

[6] A reference number in bold typeface signifies studies that use the approach, non bold signifies studies that only discuss the approach.

panels to assess research quality and impact based on an assembled evidence portfolio.	also retaining panel-specific criteria to ensure rigour and relevance • Can be adapted for different needs (e.g. the Irish Environmental Protection Agency's adaptation of the framework, which is less resource intensive but does not include independent individuals on the expert panel)	and aid to learning • Independence of the evaluation determined by composition of the expert evaluation panel
RAPID outcome assessment: **Total 5: 71**, 72, **87**, 156S, 158S A participatory mixed-methods framework to map visually the impact of a programme (including research, events and actors) on policy and the policy environment.	• Comprehensively maps cause and effect of key events • Assesses contribution of key actors • Mixed methods to counteract bias • Has been tried and tested in the international development field	• Costly and time-intensive • Does not capture the economic impacts of a programme.
Discrete research impact evaluation tools; these include individual methods that can be incorporated into certain stages of the broader impact evaluation framework e.g. to explore potential areas of impact, to put an economic value on previously identified impacts, or to address the issue of attribution etc.		
Impact pathway mapping/logic models: **Total 6**: **153S**, 154S, **161S,** 196S, 201S, 202S Involves the construction of hypothesised impact pathways, mapping out assumptions of where the information from research entered the policy-making system, the paths it tool, and the decision-makers it influenced. These assumptions of transmission and uptake can then be explored or verified using multiple methods, including, for example, key informant interviews, and primary/secondary data sources.	• Allows the pathways of impact to be hypothesised and data needs to explore and verify them to be specified • Can map routes towards impact, as well as actual impact • Can highlight where research influence occurs but does not lead to tangible policy change (e.g. research findings used in debate but does not lead to a policy reform) • Can help highlight unexpected/unpredicted impacts	• Risk that large domains of the logic model are missing or poorly defined • No formal statistical tests yet suitable for conducting significance tests of how the observed impacts match the hypothesised or intended impacts i.e. hard to assess causality/attribution • Can be labour/time-intensive
Episode studies: **Total 4:** 26c, 71, 72, 87	• Both direct and indirect observations • Assesses relative importance of influences • Process of working backwards can capture complex policy processes	• Over-emphasises political factors and under-emphasises the role of research • Tracks backwards to consider factors influencing a particular policy change, so not as useful for

Uses a narrative to tell the history of events around a policy change. Uses a four element framework (political context, use of evidence, links and relationships, external factors) to identify factors, key actors and events that have contributed to a policy change. Mixed methods used, including interviews and documentary analysis.		organisations wishing to track forwards from their research • Risk of actors 're-writing' history
Simulation: **Total 3**: 41, 48, 79 The project/programme goals and constraints on meeting them are modelled using numerical techniques to predict the impact of policy. Uses probability-based operations research modelling approach.	• Can reproduce variation and predict impact of multiple objectives • Can help choose between different options • Can be conducted *ex ante* or *ex post*.	• Complexity of model determines cost • Limited by the extent of understanding of the interaction between research and policy • Hypothetical, limited real world data
Economic analysis (e.g. cost-benefit analysis, 'willingness-to-pay' techniques, partial equilibrium models): **Total 28:** 15, 28, 48, **51**, 76, 79, 79a, 80, 82, 90, **105, 107**, 113, **124**, 128, 132. 137, 142, 146, 151S, 152S, **153S,** 154S, 155S, 170S, 196S, 197S, 208S These approaches involve the analysis of inputs, outputs and outcomes quantified in economic terms and are often used to evaluate the effectiveness of research programmes.	• Enables comparison of programmes on the basis of pre-established criteria • Discounted rates of return can account for time lags and weight factors according to importance • Easily understood • Can help determine what benefits additional research funding will yield	• Difficulties in quantifying outcomes, social welfare and externalities • Subjective interpretation of attribution and what to 'cost' • Resource-intensive method • Time separating research from economic benefit is often long • More applicable to applied than basic research; existing methods and data are sufficient to measure only a subset of important dimensions of the outcomes and impacts of fundamental science • Comparisons of cost-effectiveness of projects can be difficult with different types of output/outcome.
Social analysis: **Total 4**: 15, 48, 76, 79 Used to value the benefits of research according to quality of life/social rates of return relative to research	• Enables comparison of programmes on the basis of pre-established criteria • Discounted rates of return can account for time lags and weight factors according to importance	• Not all outcomes are quantifiable – requires subjective judgement • Measures impact of policy rather than impact on policy

inputs.		
Use of a counterfactual: **Total 5**: 151S, 152S, **153S,** 154S, 197S Counterfactual scenarios can be generated to understand what wider benefits of a particular policy would likely have been generated with all other players in the process still active, but with the research project in question removed. This helps to identify what proportion of the wider benefits of a particular policy may be attributed to the research.	• Is regarded as a rigorous approach for trying to overcome the attribution issue	• Can be time/labour intensive • Difficulties in orienting interview and survey data around counterfactual identification
Network mapping and analysis **Total 7:** 05, 48, 82, **88,** 151S, **153S,** 170S Analysis of the structure of relationships and the consequences for actors' decisions on actions. Mapping can identify multiple links (structure) and surveys/interviews can explore how they are used and valued (agency). Can be examined from the perspective of a single actor or total network (formalised survey, complemented by interviews, workshops etc).	• Reflects complex interactions of the realities of policy-making	• Can identify linkages but cannot identify impacts/outcomes as a consequence
Positive utilisation narratives **Total 4**: 01, 39, 72, **127** A participatory method to identify stakeholder accounts of impact, and the use of secondary analysis to understand who contributed to change and how.	• Identifies unexpected change • Reflects organisational values • Quick and cost-effective method of evaluation • Self evaluation and learning	• Anecdotal accounts and within-organisation evaluation prone to bias • Does not identify negative impacts or the non-use of research
Tracing post-research activity **Total 1***:* 101	• Broadens the scope of research 'user' • Seeks to understand the dynamics of research flow, and interactions between key actors	• Difficulty in tracking activities • Anecdotal

Follows the impact of research according to channels of diffusion through networks and post-research activity of researchers.		

Historical tracing **Total 7**: 15, 18, **45,** 73, 79a, 80, 170S Tracing backwards from an outcome to identify contributing factors, using a range of (usually qualitative) data collection methods.	• Policy oriented approach to tracking events and processes • Can be used to explain 'how', 'what' and 'why'	• Relies upon the quality of, and access to, existing documentation • Difficult to attribute causality • Does not account for indirect impacts, including dead-ends, spill-overs, and synergistic effects.
Retrospective analysis **Total 3:** 155S, 196S, 197S This approach tries to reconstruct history, focusing on multiple scientific innovations/policy changes. The goal is to identify linkages between innovations/policy change and particular types of antecedent events (usually either funding or research). Usually done by a panel of experts.	• Able to capture outcomes of research across many decades • Illustrates the linkages and factors necessary to advance research • Useful as a tool for distinguishing 'policy churn' (changes in the policy environment that may influence impacts) from economic impact i.e. some impacts last through changes of government or policies and others do not	• Not as useful as a short-term tool because of the long interval between research and policy outcomes • Incorrect time interval may miss important factors or events • Dependent on human factors such as marketing of research and interest by professional audiences • Difficult to capture many contributing factors
Indicators and metrics-based approaches; these can be incorporated as one aspect of a wider research impact evaluation e.g. to provide a numerical rating of impact against set criteria so that comparisons can be drawn across organisations or between the perceptions of different individuals involved in the evaluation process.		
Benchmarking: **Total 10**: 15, 46, 48, 79a, 112, 116, 151S, 155S, **178S,** 208S Uses baseline indicators against which to monitor progress towards objectives. Useful methodology for ongoing monitoring.	• Measures specific pre-defined outcomes • Provides a potential tool for comparison across programs and countries	• Objectives not necessarily measurable • Does not attribute causality • Policy has multiple-objectives; no single aggregate indicator appropriate • Can provide incentives to achieve • Very sensitive to the chosen indicators and may not take proper account of differing contextual conditions
Balanced Scorecard:	• Focuses on direct attribution of a programme as it	• Short term impact – less useful for assessing

Total 1: 121 A ten-step framework to evaluate research programmes /organisations across four perspectives; employee learning, internal business, financial, and client relationships. Impact is just one of the categories evaluated.	tracks immediate impacts • Performance indicators reflect specific goals and objectives • Learning, accountability and strategy	attribution of longer term outcomes • Useful for evaluating programme impacts, not necessarily impact on policy
Bibliometrics/citation analysis: **Total 47:** 09, 11, 15, 18, 21, 23, 28, 29, 30, 34, **36**, 41, 52, 56, 61, 64, 65, 72, 73, 76, 78, 79, 79a, 80, 81, 82, 97, **96a,** 98, **104**, 106, 127, 131, **133, 134**, 137, **143, 144**, 151S, 152S, 155S, **169S**, 170S, 179S, 180S, 196S, **202S** Method for quantifying the impact of research by counting the number of outputs and citations, and analysing citation data.	• Suitable for repeated analyses and comparisons • Measures original research, not the programmes resulting from it • Identifies research deemed to be important by subsequent research. • Traditionally considers academic citations, but extended analysis includes grey literature • Data easily available • Relatively inexpensive • Provides rapid results • Correlates with other evaluation methods	• Measures outputs not outcomes; citations by researchers only represent the first of many steps by which research is put to use by society • Not comparable between disciplines • Quantity of output may not reflect quality; risks distorting the importance of select publications • Limited role where publication is not the goal of research • Many uncertainties, and thus can only be used as a partial indicator • Time lag between publication and citation • Can be artificially influenced
Patents/new technologies **Total 6:** 11, 18, 34, 41, 76, 82 Where research may have patentable outcomes, this approach gathers data about the number and nature of patents.	• Useful to identify linkages between research and specific outcomes.	• Difficult to compare between disciplines • Time lag after publication • Most commonly used to evaluate the wider impact of research on industry.
Tools that focus on *ex ante* research management and objective setting. These tools enable the evaluation of impact through a focus on the research process and its management, for example, through formalising the impact objectives of a research programme at an early stage of research. The impact objectives can then be evaluated either during the research or following research completion.		
Logical Framework Analysis: **Total 3:** 49, **117,** 151S	• Evaluation is embedded in the planning process	• Does not integrate contextual analysis • Log frame does not represent reality

Logical framework analysis is a planning and management tool which aims to facilitate clear and specific thinking about what a policy programme aims to achieve, and how, and highlights factors which the achievement of these objectives depends on. The framework is 'logical' as it makes explicit the links between a programmes activities, outputs, purpose and goals and the assumptions that these links are based on i.e. what assumptions need to be true for 'activities' to deliver 'outputs' and for 'outputs' to deliver the 'purpose' of the programme etc. Data are presented in a matrix and mapped against indicators of achievement, means of verification and risks and assumptions.		
ROAMEF: **Total 3**: 37, 56, 118 This framework is recommended in the UK Treasury's Green Book on evaluation, and reflects the key stages of the policy process; **R**ationale, **O**bjectives, **A**ppraisal, **M**onitoring, **E**valuation, and **F**eedback.	• Puts evaluation within a wider framework describing the policy process	• Does not specify methods for conducting evaluation
Results-Based Management and Accountability Framework: **Total 2:** 116, 121 Framework to monitor outcomes through the process of activity; results chain, resources, process, outputs, client reach, outcomes, and impact.	• Focuses on attribution by guidance through the 'results chain' • Helps set target and track progress	• No prescribed methodology • Linear model – not reality • Useful for evaluating programme impacts, not necessarily impact on policy
Outcome mapping: **Total 8:** 42, 43, 52, 71, 72, 87, 116, 127	• Assesses both process and outcome • Assesses contribution (not attribution) • Enables stakeholder dialogue, learning and consensus building, as well as accountability	• Does not evaluate value for money • Observed behavioural change posits a link between cause and effect which may be impossible to demonstrate

Participatory planning tool to identify and monitor progress towards intended outcomes (defined as changes of behaviour, relationships, activities or actions of boundary organisations) i.e. both ex-ante and ex-post elements to the approach.		• Costly and time-intensive • Captures elements of policy implementation rather than focusing on ‘paper’ policy

3.5 Five Important Evaluation Frameworks

As illustrated in Table 4, five evaluation ***frameworks*** were identified by this and the previous Defra review:

- HERG Payback Model,
- Research Impact Framework,
- Irish Environmental Protection Agency Framework,
- The Australian Research Quality Framework,
- RAPID Outcome Assessment.

These methods are an important part of the Guide and Guidelines and this section gives some further background material on them.

3.5.1 The Health Economics Research Group (HERG) Payback Model

The Health Economics Research Group (HERG) Payback Model consists of two main conceptual elements. Firstly, it has five 'payback' categories that are the areas of interest in which impacts may have occurred due to research. Secondly, it has a payback model which indicates the relationship between the payback categories and the mechanisms by which paybacks occur, i.e. a model of the research process from pre-project to post-project and ultimate impacts. The payback model provides a standard framework for the capturing and analysis of research impacts and mechanisms of impact from multiple projects.

The approach was developed by the Health Services Research Group at Brunel University in order to examine the impact of health services research, but has recently been applied to other areas such as basic and biomedical research and social science research. This framework tends to employ a mixed methods approach, including both quantitative (e.g. bibliometric analysis including 'grey' literature) and qualitative methods (case studies, documentary analysis, interviews with key stakeholders, questionnaires).

The following summary is taken from Hanney *et al.*, 2000[7].

The HERG approach categorises the types of 'payback' (benefits from research) in five domains, most of which contain further subdivisions:

1) ***Knowledge benefits;*** the contribution to knowledge, be it new, confirmatory or just local evidence of something already widely acknowledged in world literature. This has traditionally been measured using peer review processes, but these can usefully be supplemented by bibliographic techniques and patent analysis.
2) ***Benefits to future research and research use;*** the place of these benefits in the model can be complex. Subdivisions may include: better targeting of future research; the

[7] Hanney S, Packwood T and Buxton M (2000), *Evaluating the Benefits from Health Research and Development Centres, A Categorization, a Model and Examples of Application*, Evaluation, Vol. 6, No. 2, 137-160, available from http://evi.sagepub.com/cgi/content/abstract/6/2/137.

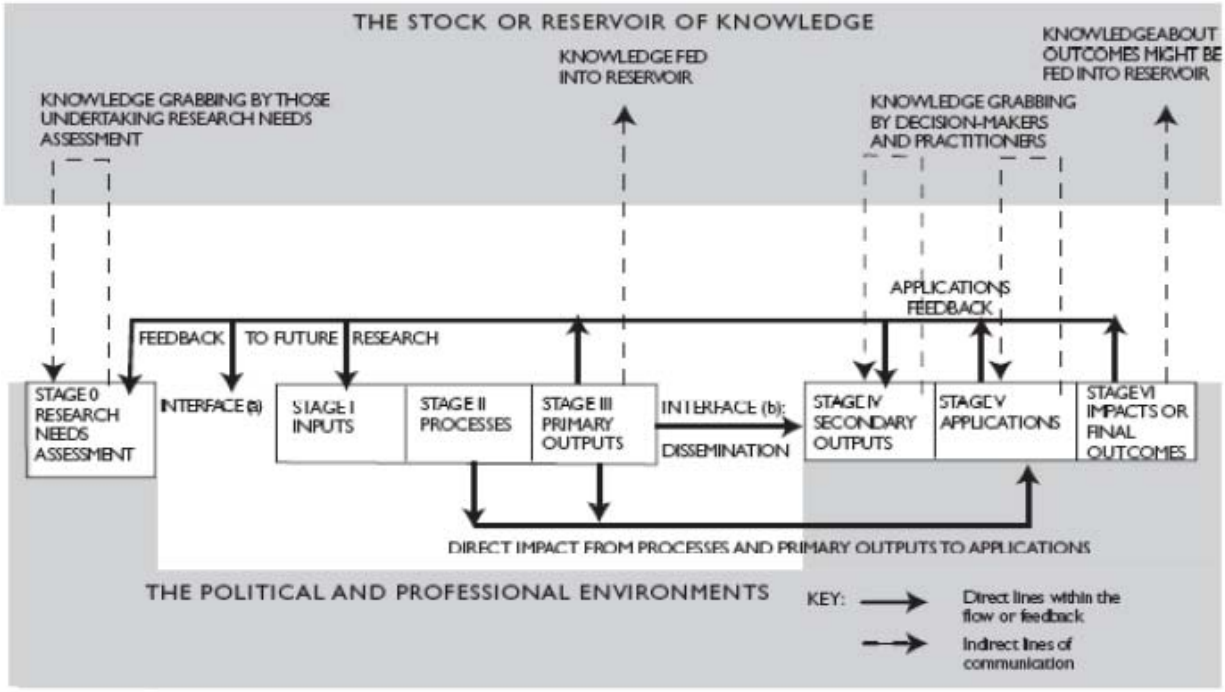
development of research skills, personnel and overall research capacity; a critical capability to utilise appropriately existing research, including that from overseas; and, staff development/educational benefits.

3) ***Informing policy and product development;*** these benefits include an improved information base and also evidence that decisions (e.g. about national or local policies, guidelines etc), were influenced by the improved information base. Other political benefits may include the use of research to: deflect criticism by showing the problem is being investigated; delay immediate decision-making; justify decisions taken for other reasons.
4) ***Health (and health sector) benefits;*** these benefits may be considered the 'real' paybacks from the perspective of those running health services and those who need to be convinced of the value of 'diverting' resources from provision of services to R&D. They could easily be adapted to explore *environmental* rather than *health* sector benefits.
5) ***Broader economic benefits;*** these include benefits arising from commercial exploitation of innovations arising from R&D.

It can be seen that the HERG approach considers a broad set of impacts of research which are much broader than policy impacts.

Having created the multidimensional categorization of benefits, a model for assigning payback can be developed, as illustrated in Figure 10.

Figure 10: The HERG Payback Model
Ref: Hanney et al., 2000: p146



The first component of the model in Figure 10, **Stage 0 – Research Needs Assessment,** includes activities completed to shape and define the research prior to its commencement. The inclusion of **interface (a) – project specification, selection and commissioning** - highlights an interactive approach in which researchers assist customers in defining their needs, and in which expected dissemination and use of the research are negotiated from the beginning.

At **Stage I – Inputs,** it is important to consider financial inputs, experience of the researchers, the knowledge base to which they are contributing and the opportunity costs of their involvement. Interaction between customers and contractors is important at this stage, and even more so in **Stage II – processes.** Some outputs and paybacks may flow directly from the processes involved in undertaking the research before formal reporting/publication of findings because, for example, the behaviour of practitioners could change as a result of being involved.

At **Stage III – primary outputs**, knowledge benefits are prevalent, though these depend on the rigour of the science in many cases and can be complex. Better targeting of research, for example, requires a feedback loop to needs assessment of future research, and the training of researchers not only helps encourage the uptake of research findings but also feeds back as an input into future research. Some publications from Stage III will more appropriately be seen as part of **Interface (b) – dissemination**. Dissemination activities can be seen as part of the research flow but generally the findings are less likely to go directly to policy-makers, instead entering a much broader and looser stock or **reservoir of knowledge**. If decision-makers or practitioners are influenced by reading reports etc sent to them as part of the dissemination process, this may be considered as part of the flow. In contrast, where knowledge enters the reservoir of knowledge and is picked up for later application (be it in policy decisions or in research needs assessments etc), this is better considered as 'knowledge grabbing'. The more research findings from one project mix with those from others and are later grabbed by policy-makers or practitioners, the harder it becomes to isolate and identify the benefits from any particular project.

Many of the paybacks occurring at **Stage IV – secondary outputs** can be classified as political and administrative. A decision in line with research could have been taken for a range of reasons, including: the substance of the findings influenced decision-makers; the findings enabled policy-makers to justify a decision they wanted to take for other reasons; the decision would have been made in the absence of the findings, but nevertheless they confirmed its legitimacy. Even where a policy decision is not totally consistent with research findings, they may still have played a major role in the policy deliberations.

Changes in the behaviour of practitioners/policy-makers have to occur before most of the final two categories of payback – health service and economic benefits – will be realised in **Stage V – applications**. At this stage, changes in behaviour have to be recorded, and efforts made to identify how far any changes are a product of the research. Finally, at **Stage VI – impacts or final outcomes**, the major assessment of service and economic paybacks can be made using various measurements such as calculations of cost savings, satisfaction with the

services provided etc. Even where observational measurements are possible, it remains important to demonstrate the link with the research project through attributable changes in use, against an assumed counter-factual.

Of the five frameworks identified, the HERG Payback model has been most widely tried and tested in different countries and policy contexts, as illustrated in Table 5. In each of these applications, data has been collected in a variety of ways, which illustrates the difference we have highlighted in this report between data collection methods and frameworks for structuring data and analysis.

Table 5. Developments of the HERG Payback Model

CAHS (2009) (152S)
Discusses the work conducted for the Canadian Academy of Health Sciences (CAHS) to develop an indicator based adaption of the HERG model. This was proposed in 2009 and has not yet been used, so its practicality is still to be tested. The approach proposed aims to define the impacts of Canadian health research, to learn how to improve the returns on investments in health research, to address gaps in the Canadian Institutes of Health Research framework, and to help clarify the 'attribution issue of health research'. The key question asked: is there a 'best way' to evaluate the impacts of health research in Canada, and are there 'best metrics' that could be used to assess those impacts (or improve them?). It is designed to be used as a roadmap to track health-research impacts in five categories; advancing knowledge, building capacity, informing decision-making, health impacts, and broad socio-economic impacts. Within those categories, it suggests a large number of possible indicators and metrics (numeric, descriptive and qualitative) covering a wide range of impacts. Interestingly from the SKEP project's perspective is that CAHS are trying to create a framework that can be used in a non-prescriptive way by all the bodies who might want to evaluate the impact of their health research. Highlights that health research generates a wide range of outputs with diverse outcomes. The uptake of that research is further influenced by a number of factors and demonstrating causal pathways that may lead from the research to impacts is consequently very difficult. It is therefore suggested that the problem of complexity creates the need for a standardised solution, involving the classification and mapping of the output-uptake-outcome relationships within an evaluation framework. The framework then allows identification of best indicators within various categories of impact. Using a standardised evaluation framework allows both comparison of evaluations and the identification of unexpected outcomes (as a framework can help to ensure that all possible outcomes are investigated).
Kalucy *et al.*, (2007) (169S)
Concludes that the HERG Payback framework is a feasible means to assess impact, but suggests minor adjustments to, and clarification of, some impact categories. It identifies some gaps e.g. relating to impacts in terms of collaborative links with the community and decision-makers to facilitate evidence-based practice and policy. Concerns are expressed about relying on reporting by investigators, suggesting possible bias in terms of: recall; over or under estimation of the effect of the research outcomes; and

measurement error. The study team experienced some respondent fatigue resulting from the length of the detailed questionnaire which might also lead to response and information bias. The paper highlights the issue that researchers in different traditions or cultures vary in the way they conceptualise and explain the impact of their research.

Nason *et al.*, (2008) (185S)

Applies the HERG model to the impacts of eight Health Research Board (HRB) grants funded in Ireland. Impacts are identified using a case-study approach, built on the HERG payback framework. The framework used consists of two elements: a logic model of the research process that tells the story of each case study, and a multi-dimensional categorisation of the benefits from research. Concludes that the payback model provides an effective framework to evaluate research, but that this could also provide a starting point for a more quantitative assessment of the economic return of identified impacts of HRB-funded research.

Suggests that while these case studies are valuable in providing a deep understanding, a complementary approach could be to develop a 'light touch' payback questionnaire that could be incorporated into end-of-grant reports to provide an overview of the impacts emerging from the entire portfolio of HRB-funded research. The questionnaire could also help to identify projects for a more detailed case study based evaluation.

Ootwijn *et al.*, (2008) (191S)

Assesses the impact of individual health technology assessment (HTA) projects by adapting the 'payback framework'. Dossier reviews were conducted and a survey sent to principal investigators of forty-three projects. The resulting documentation on outputs and outcomes was summarised and assessed by ten experts using a scoring method (which the authors indicate needs to be further developed). Five case studies were compiled using information from additional dossier review and semi-structured key informant interviews.

Concluded that the payback framework is a useful approach to assess the impact of HTA projects. The case studies provided greater depth and understanding about impacts and why and how they have been achieved. The study concluded it is too early to determine whether the HTA programme led to actual changes in healthcare policy and practice but that the results from it can be used as a baseline measurement for future evaluation and considerations of how to assess impact.

Wooding *et al.*, (2005) (207S)

Applies the HERG model to 16 case studies representing different types of research grants funded by a single body and awarded to projects, programmes, fellowships, an institute, basic research, clinical research and health professionals. Evaluation data was collected from document and literature reviews, semi-structured key informant interviews and bibliometrics. The case studies were written up in the framework provided by the payback model and the cases were compared – which was facilitated by their being in a common framework. Two approaches were used to assess the cases. The first was qualitative assessment and the second a novel scoring method. The approach involved a two day deliberative workshop and adjustment of the scores by the participants.

3.5.2 The Research Impact Framework (RIF)

The Research Impact Framework (RIF) is an easy-to-use and quick tool that includes descriptive impact categories to prompt researchers to systematically explore and describe the impact of their work. It was developed by the London School of Hygiene and Tropical Medicine in order to guide an analysis of the impact of a selection of their research projects.

The RIF is not itself intended to be evaluative in terms of prioritising impacts or proposing causal pathways. It seeks to: act as a standardised framework to help capture impacts across research topics and methods and facilitate comparison across time and cases; to guide researchers in planning research implementation and evaluation strategies; to assist researchers in looking at the broader influences and effects on and of their work in society; to promote research accountability in relation to the use of resources and the consequences of research; to help in the attribution of effects to health research given the range of other determinants of health and societal impacts; and to contribute to more extensive or specialised evaluations of research impact (Kuruvilla *et al.*, 2006).

The RIF is primarily a narrative-based, descriptive approach, and is a four-dimensional framework providing an overview of potential (sometimes overlapping) research impact areas. The four narrative areas of impact captured are:

1) *Research-related impacts* e.g. research methods used, publications and papers, research networks, research management, dissemination and communication etc.
2) *Service impacts* e.g. evidence-based practice, quality of health care etc.
3) *Societal impacts* e.g. health literacy, health status, knowledge, attitudes and behaviour, social capital, sustainable development outcomes etc
4) *Policy impacts* e.g. level of policy-making influenced (local, regional, national etc), type of policy (e.g. practice, service, governance policies), the nature of policy impact (i.e. conceptual, instrumental etc), policy networks (the extent to which researchers are part of, or inform, policy networks), and political capital (the value of research evidence and researchers themselves in policy negotiations, in reaching high quality agreements and improvements to the policy-making process etc.).

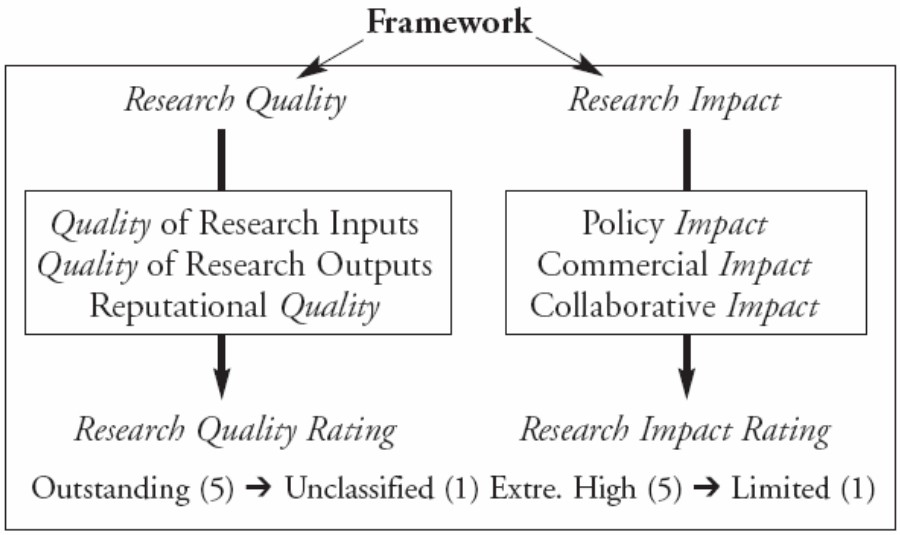
Themes within each narrative area can be removed, added to, grouped, or modified as appropriate to the research being described. It is emphasised that researchers should be encouraged to think about the negative as well as the positive impacts of their work, such as stigma that could arise with publication of research findings related to a particular group or community.

3.5.3 The Irish Environmental Protection Agency (EPA) framework and the Australian Research Quality Framework.

The Irish EPA framework and Australian Research Quality Framework are closely related approaches. The Irish EPA framework is primarily an indicator-based approach, with indicators sitting under an umbrella of two main criteria: research quality and impact (see Figure 11). Each criterion is divided into three sub-criteria (each with several indicators) and data is collected on each of these indicators and used to rate both the quality and impact of the research in a simple numerical format. The three sub-criteria of research impact are:

1. *Commercial impacts*: the effects of the research on commercial activities such as patent applications or savings in resource use;
2. *Collaborative impacts:* this measures the linkages between the researchers and the policy community, commercial enterprises and other key stakeholders;
3. *Policy impacts:* this measures the contribution of the research to national and international policy development including work on advisory panels to industry or government agencies.

Figure 11: Irish EPA Research Impact Framework
Ref: O'Leary et al (2008): p162



The data required to assess the research against each of the sub-criteria is collected using a range of methods such as documentary analysis, citation analysis, interviews with project personnel and/or end-of-project questionnaires sent to project co-ordinators. A context statement is also written for each project and this allows the opportunity for other factors not captured by the indicators to be factored into the overall rating. This aspect builds in some element of flexibility to the fairly rigid framework.

Based on the data collected, both the quality and impact of the research are rated by an in-house panel of experts, using a five-point rating scale. This rating is used to produce two scores for the research; one reflecting quality and the other reflecting impact. Subsequent use of this approach in the Environmental Technologies Research Programme also included the production of a series of case studies alongside the final impact and quality scores. These case studies were based on the responses to surveys sent to the principal investigators of 18 projects within the programme, asking about a number of payback categories, including: knowledge production, capacity development, informing policy and environmental benefits, and broad social and economic benefits (EPA, 2009[8]). The Irish EPA approach flags up the potential for incorporating research quality into a theoretical framework.

The approach was largely derived from the Australian Research Quality Framework (RQF), which is more comprehensive but was abandoned prior to its implementation following Australia's 2007 election.

The key difference between the two approaches relates to the expert panel composition and the evidence used to inform its decision. The evidence portfolio required in the Australian RQF included: the context statement; four 'best' research outputs per researcher; a full list of research outputs; statements of impact that could be verified by qualified end-users of research; and other discipline-specific measures. The impact statement included: an evidence-based statement of claims for the Research Group against generic and panel-specific impact criteria, including verifiable indicators in support of these claims; up to four case studies that illustrate the Group's claims of impact; and, details of end-users who can be contacted by Assessment Panels to verify the Research Group's claims.

Both the Irish and Australian approaches use an expert panel to assess quality and impact based on the evidence portfolio. However, the Australian panel was proposed to be made up of 12 people, including a mixture of end users and internationally renowned peers, i.e. is independent. In contrast, the Irish EPA one consists of two internal and one external expert (this is a practical response to the scale of evaluation and the resources available). Following completion of the review we have been made aware of efforts being made to engage a greater number of stakeholders in the Irish EPA evaluation process, primarily through consultation in the form of extensive meetings, emails and telephone communications with the research-project leaders, relevant EPA personnel and other key project stakeholders. These are documented in Wemaere *et al.* (2009)[9].

While both approaches use context statements, the Australian RQF context statement is also accompanied by an in-depth impact statement, which includes verifiable, evidence-based claims against specific impact criteria, up to four case studies that illustrate examples of those

[8] Irish Environmental Protection Agency (2009) *Innovation for a Green Economy. Environment and technology: a win-win story, Irish EPA.* Available at www.epa.ie/downloads/pubs/research/tech/EPA%20STRIVE%20ET%20Report_Final.pdf

[9] Wemaere A, Kilroy G, Sheils, L and Donlon, B (2009) An evaluation of the role of EPA research in the Water Framework Directive implementation in Ireland. *Biology and Environment: Proceedings of the Royal Irish Academy* 109B, 385–402.

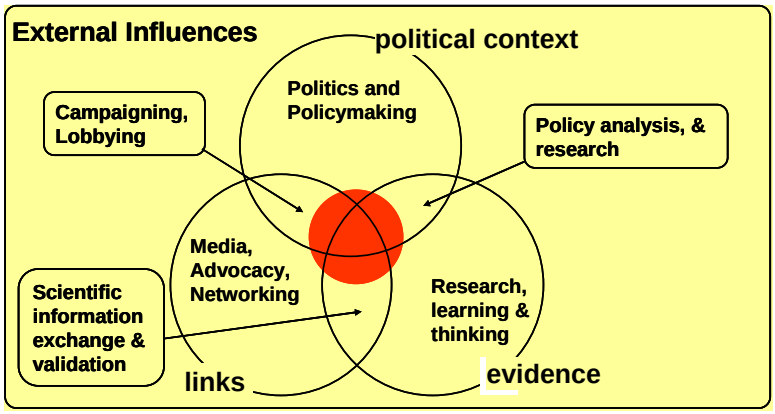
claims, and details of end-users who can be contacted as referees to verify the claims, again highlighting the more objective nature of the Australian approach.

3.5.4 RAPID Outcome Assessment (ROA)

RAPID Outcome Assessment (ROA) is a learning methodology to assess and map the contribution of a project's actions to a particular change in policy or the policy environment. It is a flexible and visual mapping tool which can be used alongside other evaluation tools and methods. The framework was developed by the CGIAR-affiliated International Livestock Research Institute (ILRI) and the Overseas Development Institute's (ODI) Research and Policy in Development (RAPID) programme, in response to the need to better understand the processes and mechanisms that lead to pro-poor decisions at the policy level. It combines elements from the International Development Research Centre (IDRC) Outcome Mapping approach (examines the progressive behaviour changes amongst the key actors that the project is directly influencing), episode studies (tracks back from a policy change to identify the factors that contributed to it), case study analysis (tracking forwards from research), and Most Significant Change (helps to identify and prioritise the key changes). ROA is designed to assess the contribution of a project's actions and research on a particular change in policy or policy environment, both during and after a project is completed.

The method involves triangulation and integration of data collected using three different methods including classical case study (what was done, what happened), behavioural changes among key actors during the life of the policy process, and an examination of the factors contributing to that change interpreted within the ODI's *Context-Evidence-Links* framework (see Figure 12). This framework considers the process and influences from the perspectives of political context, role and use of evidence, role of linkages, and effect of the external environment.

Figure 12: Context-Evidence-Links framework
Ref: *www.odi.org.uk/RAPID/Tools/Toolkits/RAPID_Framework.html*



The method follows a three stage approach:

1. The first stage is a preparation stage, in which a documentary review and a series of informal conversations are carried out to develop an initial picture of the project's history and the intended changes. This will be guided by the Context-Evidence-Links framework, and includes case study (or project-based) information and episode study (or policy-change focused) information.
2. The second stage is a participatory workshop, in which participants (typically including the project team, key stakeholders and other external experts) identify the key policy change processes and causes of change in order to build a map of influences. The workshop covers the following:
    - Defining the policy environment at the start and end of the project/period.
    - Identifying key policy actors and 'boundary partners' that were considered influential in the process of ensuring impact. These may be individuals or institutions, and can be clustered into groups, for instance 'Civil society organisations', 'donors', 'the private sector' etc.
    - Describing the behaviour of the key actors/boundary partners that contributed to the change in the policy environment or policy, and the behaviour at the beginning of the project.
    - Establishing a time-line.
    - Mapping key behaviour changes for each of the key actors/boundary partners along the timeline, from the beginning to the end.
    - Mapping both the key changes in the project (including organisational changes, outputs and changes in behaviour), and the external influences including actions of strategic partners and other exogenous partners, during the same period.
    - Determining the level of impact/influence of (i) the project on the changes in behaviour of the key actors/partners, and (ii) of external influences on the changes in behaviour of the key actors/partners and the project i.e. exploring the links between the actors' behaviour changes and the identified events, including project activities, external influences and other actors' behaviour change.
3. The final stage involves a follow-up 'de-briefing' process allowing researchers to develop and refine stories of change – a 'timeline' - identifying key policy actors and events and their contribution to change. The timeline is used to identify key informants to follow up with in-depth interviews, which aim to help confirm the linkages and influences determined in the workshop and to assess the nature of the contributions to change.

3.6 Complementary evaluation tools and metrics

Table 4 highlights the wide range of evaluation tools, indicators and metrics identified that have been used in past impact evaluations, often as part of the five main evaluation frameworks discussed in the previous section. They fall into three broad categories:

1. Research impact evaluation tools
2. Indicators and metrics
3. Tools for research management and objective setting.

Three of these evaluation tools are frequently reported to have been used within the literature:

- Economic approaches to evaluating the impact of policy-oriented research
- Logic models
- Use of a counterfactual

Details of these are given below. Details of some of the other tools, indicators and management tools are given above in the section on the main frameworks and in the case studies.

3.6.1 Economic approaches to evaluating the impact of policy-orientated research

Economic approaches are routinely used in the evaluation and appraisal of policy options to provide policy-makers with quantified estimates of the costs and benefits of different policy options. There is a wide literature on the possible approaches that can be used and their strengths and weaknesses.

Molas-Gallart and Tang (208S) report on a workshop organised by the UK's Economic and Social Research Council (ESRC) that discussed the application of economic approaches to the assessment of the impact of social and economic research.

The paper provides useful summaries of four case studies intended to understand the impact of various ESRC research activities which are highlighted elsewhere in this report. It also addresses the question of how economic approaches could have built on these evaluations, particularly in the context of the policy expectation in the UK that research funders should justify the value gained from investment of public funds. The workshop considered economic approaches that could be applied to research impact evaluation including: willingness to pay; case study approaches; quantitative approaches such as econometrics and economic modelling; peer review; and international benchmarking. It concluded that all the approaches could in principle be used to demonstrate accountability of research funding. However, to learn about the impact of research on previous policy efforts, only case studies (as a basis for developing quantitative estimates of impact) and, to a limited extent, career tracking, (using an economic indicator such as salary differentials to track impact over time) could be used. Given these results are based on workshop discussions some caution should be exercised in adopting them uncritically.

Ryan (199S) considers the impact of 'policy-orientated social science research' on policy. Impact is evaluated through a consideration of the degree to which research has advanced the speed of policy development and the economic benefits that can be attributed to this faster

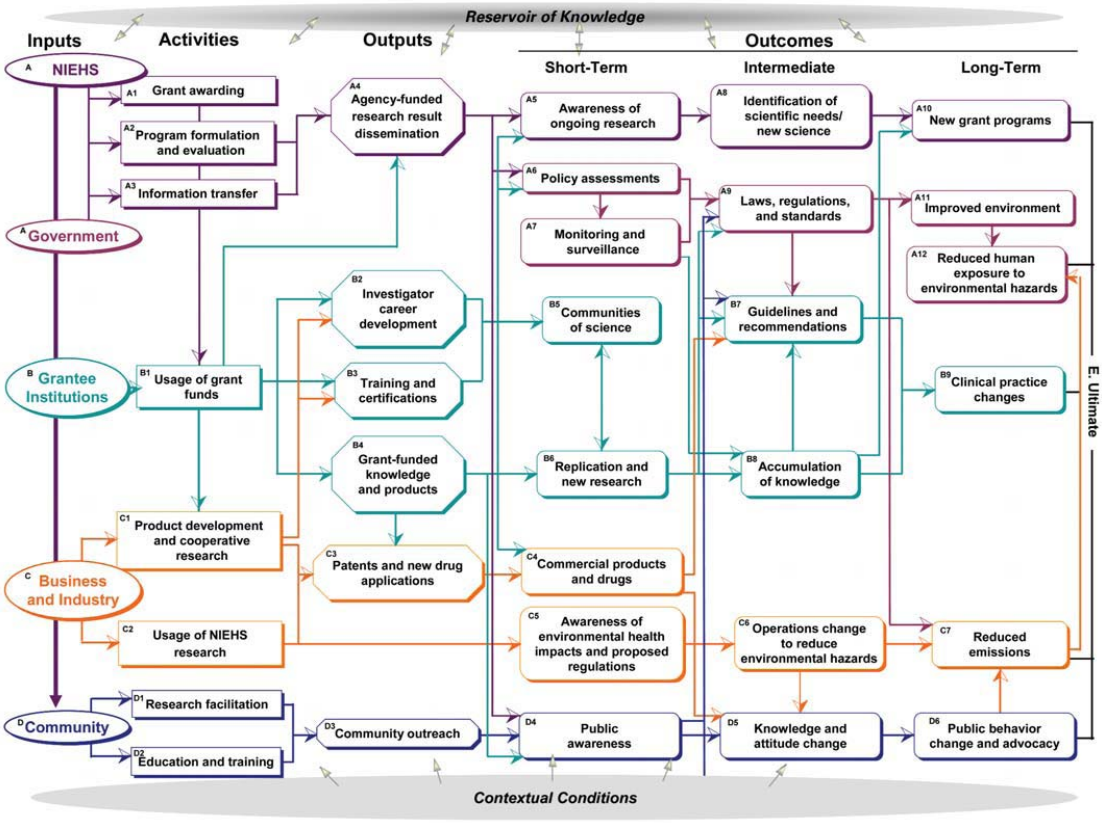
development. The approach combines qualitative interviews with the researchers and policy community and established that the research had resulted in policy decisions being taken earlier than they would have otherwise been taken (in the Vietnamese rice commodity market). The economic impact of these earlier decisions was estimated using the partial equilibrium framework economic model to estimate the potential domestic and international welfare benefits of the research. While Vietnamese rice commodity markets may seem a rather obscure area, the approach illustrates how economic measures can be used to quantify impacts and how they can be linked with qualitative methods. The approach may appear complex, yet its final outcome is in a form that is easily communicable; it is given as an economic measure and therefore the impact of the evaluation may be greater.

3.6.2 Logic models

A number of examples of logic models have been used and discussed in the literature, and the central element of the HERG model is its logic model. 202S is one discussed above and the central element of the HERG model is its logic model. However, 161S focuses exclusively on the development of a logic model for a specific evaluation purpose. It gives an example of how logic models, that is '*graphic depictions of the relationship between a program's activities and its intended outcomes*' - can be developed to assist in measuring the impact of environmental health research programs on human health, the environment, and the economy, even when the impact may be indirect or diffuse.

The approach uses expert input and literature/documentary review on research impact assessment to develop a logic model that defines multiple components and linkages between (environmental health research) grant programmes and their outputs and outcomes in a variety of areas. The model delineates pathways, in which five types of institutional partners are included in the research process: the research funder, other government agencies, grantee institutions, business and industry, and community partners. Figure 13 illustrates the final logic model.

Figure 13: Example of logic model approach to research impact evaluation
(Source: Engel-Cox et al., 161S) Arrows represent linkages between the logic model components. Pathways are identified by letter and colour. Full details in original source.



This example presents a rather complex logic model linking multiple components that go beyond the contribution of research. However, the approach can be built up from fewer elements, or by focusing on one set of linkages of interest, for example, the linkages between research and policy. Individual elements and their assembly are discussed in 161S. In essence the value of the approach is derived from its ability to specify elements of the pathways of impact and potential linkages between them. From this process the data needs to assess the mechanisms of impact and their importance are likely to become clearer and can be specified and developed.

Another example is provided by Trochim *et al* (202S). This describes and discusses lessons from a pilot attempt to evaluate a large scale research initiative (US$70 million) in a US government agency (the Transdisciplinary Tobacco Use Research Center (TTURC) initiative of the National Cancer Institute). A mixed-methods approach was adopted included concept mapping, logic modelling, a detailed researcher survey, content analysis and systematic peer-evaluation of progress reports, bibliometric analysis and peer evaluation of publications and citations, and financial expenditures analysis. Of central importance to the approach was the early development of a conceptual framework and logic model accounting for possible research impacts in the short, medium and long term and the ordering and analysis of the evaluation data collected:

> ‘The logic model is the key unifying device for organizing and grouping results from multiple methods for each outcome area and enabling synthesis of the findings.’ Page 16, 202S

For reasons of efficiency the evaluation relied wherever possible on pre-existing data sources rather than creating new measures. The only new significant data source specific to the evaluation framework was a Researcher Form – an annual survey of investigators and research staff to elicit their opinions and evaluative assessments regarding the entire range of outcomes in the logic model, including collaboration, transdisciplinary integration, science, models and methods, internal and external support and recognition, communications, and the effects of research on policy, practice and health outcomes.

The approach of logic models has been generalised in the approach of the Logical Framework Analysis (as noted in the previous Defra review). Logical Framework Analysis is a planning and management tool with four key dimensions: goals, purpose; outputs; and activities. Data are presented in a matrix and mapped against indicators of achievement, means of verification, and risks and assumptions. The two dimensions are linked by a set of logical prepositions and assumptions about the delivery of activities, outputs, purposes and goals.

3.6.3 Use of a counterfactual

The Consultative Group on International Agricultural Research (CGIAR, 2008) uses a number of evaluation tools in their approach, drawing together economic approaches, impact pathways/logic models and the use of a counterfactual (153S).

While each case is slightly different, the essence of CGIAR’s approach is the construction of hypothesised impact pathways, in which the authors charted assumptions about where information from the research entered the policy-making system, the paths it took, and the decision-makers it influenced, and then sought to verify these assumptions of transmission and uptake. In most cases, this verification process involved conducting a large number of in-depth key informant interviews to explore influence and impact, together with the use of primary and secondary data sources and reports to validate and support the information gained from these interviews (the key informant approach can be useful where there are a limited number of people with significant enough depth and breadth of knowledge to speak informatively about what happened and why). Key informants were also used to broaden the scope of later interviews and add other interviewees to the original list. One study featured a round of feedback from the key informants so that they could comment on the perceptions of their colleagues to arrive at a more reliable subjective scoring of influence.

The second stage of the evaluation process involved quantifying the magnitude and distribution of the impacts of a particular policy change, and then generating a counterfactual to calculate what wider benefits of the particular policy would likely have been generated with all the other players in the process still active, but with the research project in question

removed. This helps to identify what proportion of the wider benefits of a particular policy may be attributed to the research.

The final stage of the evaluations was to compare the measured benefits attributed to the POR with the cost of the research to calculate Net Present Value (NPV), Internal Rate of Return (IRR) and/or the Benefit-Cost Ration (BCR) i.e. to identify the returns to the research investment.

4 What are the strengths and weaknesses of different approaches to research impact evaluation?

In the previous section we highlighted the wide range of frameworks and data collection for research impact evaluation analysis. Obviously there are a large number of strengths and weaknesses associated with these and they will inform the design and selection of any evaluation approach. Table 4 gives more detail on the strengths and weaknesses of each these approaches which have been drawn from the literature studied in and Table 3 summarises the strengths and weaknesses associated with the different methods for gathering evaluation data.

These tables present a large amount of information that is useful for those interested in the specific approaches. However, it also presents a challenge in terms of how to determine which data collection method or framework of evaluation to use in a specific instance and how to pick one approach over another. The literature has little to say on the strengths and weaknesses of different approaches in direct comparison to one another.

Given the wide range of evaluation contexts, objectives and approaches, a degree of judgement and expert knowledge is inevitably required to develop an approach that balances the tensions that may exist between these strengths and weaknesses. In this section we note the specific and generic issues associated with different approaches to research impact evaluation.

The strengths and weaknesses of different methods fall into two broad categories. Firstly, there is a set of *practical* issues to be considered, for example, issues associated with the cost of an approach, the time required to conduct it and availability of appropriate expertise. Secondly, there are more fundamental issues of *evaluation practice and theory* which relate to the limits and challenges of evaluation methods which need to be considered if successful evaluations are to be conducted. These include: how to deal with attribution of policy impact to a research intervention; the effect of time-lags between research outputs and possible policy impacts; and the theories and understandings of the impact of research on knowledge and how knowledge is translated into policy.

The strengths and weaknesses highlighted in Table 3 and Table 4 are numerous, but some key practical issues to consider in selection of a research impact evaluation frameworks and data collection methods can be drawn out:

- **Scope, scale and resources** – what are the likely time, financial and other resource implications of the approach and are these appropriate to the evaluation objectives? Is the framework applicable to outcomes of interest? Are expert evaluators or other disciplinary skills required to advise, or conduct the evaluation, or can the approach be conducted in its entirety or in part by non experts?

- **Adaptability** – Is the approach prescriptive or flexible? Can the approach be adapted to the setting of interest and data availability on it, or does it require specific data inputs which might constrain its use?

- **Subjectivity** – how does the approach deal with possible subjective data inputs or methods of analysis and/or the conflicting results from different analyses either from within a particular framework or beyond it?

- **Complexity of the research/knowledge/policy interface** – how, if at all, does the framework account for the complexity of the interactions between research, knowledge and policy and the broader context that research is being conducted in and policy is being developed in? Is there an interest in understanding the impacts and/or the processes leading to them?

- **Attribution** – does/can the framework attribute impacts to research interventions? If so how does it establish the mechanism(s) of impact? Or is the focus on establishing the 'contribution' of research to impacts? I.e. is the interest in establishing links between interventions and impact and how that research has played a role in them or is there an interest in identifying the causal mechanism and significance of the intervention relative to other potential interventions?

- **Disciplinary basis** – does the approach rely on a single disciplinary basis and might this limit the scope of evaluation? Or does multidisciplinary increase the complexity of evaluation? Can it consider and compare inputs and impacts in a range of ways, for example, in social, economic and environmental terms?

- **Nature of outputs** – are the outputs from the evaluation likely to be suited to the intended audience and users and be meaningful to them? Are the outputs descriptive or explanatory? Can an appropriate balance be struck between brevity and depth of insights? Are indicators and metrics used to summarise and communicate outputs simply and are these balanced with more detailed insights from qualitative methods?

- **Comparability** – does the framework allow comparison of evaluations of different research activities – project, programmes, disciplines etc? Can evaluation outputs be compared with outputs from other frameworks now or in the future?

- **Participatory** – to what degree does the framework involve the participation and collaboration of researchers, research managers, funders and potential users in the approach?

- **Unexpected or hidden impacts** – is the focus on defined outcomes of interest or identification of unknown outcomes? I.e. is the framework likely to reveal unexpected impacts (e.g. ones in non-target audiences) or hidden impacts (e.g. the easily

overlooked impact of research that confirms existing policy positions)? Or is it assumed that impacts will occur in a specific area?

And some issues of particular importance to data collection methods are:

- **Speed and cost** – How quickly and at what cost can the data be acquired?

- **Availability** – Is primary or secondary data required? Are sources easily accessible? Is the source documentary or survey data, for example, and is it in the public domain, and/or are responses/participation from people required? Are the relevant people likely to be available or accessible and have the required time/willingness to participate? Are data available over the period of interest?

- **Reliability** – how reliable are the data collected likely to be? What biases might they introduce and how can they be overcome? Are there sampling issues that need to be addressed, e.g. are the right people being involved and can enough people be contacted to create a reliable picture(s) of research impact when users may be diverse and diffuse? Does an approach with broad coverage but less qualitative insight (e.g. bibliometrics or surveys) need to be balanced with an approach with narrower coverage but which reveals more depth, such as interviews?

- **Repeatability** – is it intended that the data be collected on a one off or a regular and ongoing basis?

5 Which methods are most effective?

The material above makes it clear that there are numerous approaches, settings and objectives for evaluation and that there is therefore a need to tailor each approach using judgement and expertise so that it is suitable for the situation in question. Most of the frameworks adopt a mixed methods approach in an attempt to deal with the weaknesses of any one method and triangulation between methods and multidisciplinary approaches are also considered important. As such answering the question ‘which methods are most effective’ cannot be done in a meaningful or at least direct way. The literature has little to say on the comparative effectiveness of different approaches. This is to be expected when different methods are being applied to very different situations making direct comparison hard or impossible. The relative performance of different approaches in similar circumstances may make an interesting topic for further research.

6 Which methods offer value for money?

The previous Defra review found little material that considered either the costs of evaluation or value for money. This review has uncovered a few sources that comment on this issue.

Both 152S and 193S highlight that research impact evaluation is expensive. However, while considered costly, the value of benefits derived from evaluation is also considered to be high. For example in 193S the role of mid term evaluation in shaping the effectiveness of the remaining programme was considered important. 152S suggests that economies of scale could be introduced through the establishment of evaluation questions of national interest.

In terms of costs 152S suggests that previous work has indicated that 'spending between 1 and 5 per cent of the research budget to evaluate outcomes is not unreasonable, and this can be used as a 'rule of thumb' (p72, 152S). The range of evaluation costs for programmes run by a Canadian research body are given as between $150,000 and $500,000. However, as the total programme costs are not indicated this prevents an assessment of value.

202S indicates the cost of a specific pilot evaluation of a large scale research initiative in the USA as between US$400,000 and US$500,000 to evaluate the first 3 years of US$70 million programme. It attributes much of this to start-up and one-off costs involved in development of the approach and suggests that an allocation of at least 1 per cent of initiative funding for such evaluation is a reasonable benchmark. The importance of pre-allocation of evaluation funds in enhancing the capacity for and timeliness of evaluation is also highlighted.

151S gives partial details of costs for the eight frameworks it considers. Some details are quoted in absolute cash terms, others in staff time requirements and others relative to programme costs. It is suggested the European Commission allocated, or planned to allocate. 0.5 per cent of its Framework research budget to evaluation in accordance with international norms and also that the DIUS framework (see Table 2) has low marginal costs as much of the data it requires has already been collected for other purposes.

These figures provide an indication of the range of costs that might be expected to associated with research impact evaluation. They are still vague but do indicate that while evaluation may be perceived to be expensive it can improve the effectiveness of research programmes.

7 What are the issues and challenges for evaluating the impact of research on policy?

A number of fundamental issues and challenges for the evaluation of research impact on policy recur through the literature. The most important appear to be those of:

- **Time** – there is an inevitable time lag between research intervention and policy impacts and depending on the length of time this can significantly hinder the identification and evaluation of impacts,
- **Subjectivity** – many data sources for research impact evaluation are likely to be subjective,
- **Attribution** – establishing (causal) links between interventions and impacts while central to research impact evaluation poses some significant challenges,
- The **nature of the relationship and exchanges between research, knowledge and policy**.

These challenges – discussed further below - go beyond the strengths and weaknesses specific to individual approaches and apply to all the research evaluation frameworks and methods used. The Canadian Academy of Health Sciences (152S), published in 2009, succinctly summarises some of the challenges facing research impact evaluation and the status of their resolution:

> 'Particular problems always arise when attempting to evaluate research impacts: attribution, the counterfactual; threats to evaluation validity; and time lags to research impacts. Since these issues consistently arise, it can be assumed that none has been fully solved.' (Page A-276, Appendix D, 152S)

Threats to evaluation validity include threats internal to the evaluation undertaken, i.e. problems that could undermine the findings of the evaluation itself; or external to the evaluation, i.e. that the evaluation itself may not be generalisable. They are largely measurement, methodological and contextual factors which have already been covered above in the discussion of the strengths and weaknesses associated with different methods and frameworks.

The counterfactual (also addressed in 151S, 153S and 154S) refers to the need to understand what would have happened in the absence of the intervention. The idea of additionality is another way of viewing the problem, i.e. what extra impact has the intervention created. 151S explores different types of additionality and highlights that, of the eight evaluation approaches considered, only three explicitly mention additionality.

While 152S suggests that the formulation of the counterfactual is a separate challenge for evaluation, it would actually seem to be part of the problem of attribution (which is discussed

below). The counterfactual can be seen as a way of testing the outcomes of attribution, i.e. if an outcome is attributed to an intervention but a counterfactual or additionality analysis suggests it would have happened anyway, or has not added anything extra, further work is required to investigate and understand the causes of the outcome.

7.1 Time-lags

The time-lag between research outputs and policy impacts is widely noted in the literature. The time-lag between research completion and wider impact can range from almost immediate to very long term. It is suggested that 50 years between research and its ultimate impact is not unusual (155S, 161S). These very long time-lags tend to be cited in the health and medical literature, but large time-lags are cited elsewhere, for example, 20 years for uptake of research in international development policy (156S). Time-lags to impact are likely to be greater and less predictable in basic compared to applied research, which has implications for both the choice of evaluation approach and its timing (155S, 170S, 187S).

Once an impact has occurred, the persistence of its impact, i.e. how long the impact lasts, will affect the ease of its evaluation and the chance of the impact being overlooked (152S, 154S).

Time-lags to, and persistence of, impacts creates problems for evaluation. Firstly, the evaluation needs to be conducted at an appropriate point to capture the impact of interest. 193S notes that evaluations are often conducted as research is completed or just after and that this can be too early for impacts to have emerged.164S suggests that evaluations should take place at different points in time to capture the different ways that research influences policy actors and the policy process over time, and 170S comments on the time-lags observed for different degrees of outcome to occur.

With increasing time and distance from the research there is the problem of declining data availability. As discussed in 161S, research influence becomes increasingly diffuse as it is traced through the logic model and suffers from time discontinuities and lack of documentation. This will be a general problem for all methods especially those based on input from actors involved in the research process or its use.

7.2 Subjectivity

Many of the sources used for research impact evaluation rely on the memory and perceptions of those involved in conducting research or users of it. This poses problems for rigorous analysis that are highlighted widely in the literature. 154S for example, suggests greater reliance needs to be placed on the use of secondary data analysis, of control sampling of non-beneficiaries, and of longitudinal surveys to get a more reliable idea of what might have happened without the intervention and its consequences. The CGIAR study (153S) also suggests corroborating views of interviewees by consulting with other interviewees.

158S also suggests care should be taken in basing impact evaluation purely on what people remember: People may have their own reasons for remembering the past in ways that suit their image of the present. People are also notoriously selective, partial, biased and prone to leave out what they assume can be taken for granted.

197S reviews and assesses the CGIAR POR impact assessment studies conducted to date. It indicates that many conclusions about policy influence are qualitatively inferred; interviews and statements about the role of research are applied in these studies to establish narratives that link the evolution of findings with policy outcomes. Quantitative metrics of influence are rare, and in some cases the evidence behind the narrative is not clear.

The response to this problem is largely suggested to be the adoption of mixed methods approaches that triangulate between the analyses resulting from different methods. 158S proposes a number of strategies to overcome this problem:

- Finding out about the assumptions, interests and pressures affecting each informant as well as the researchers. To test conflicting narratives and cross-check findings, various interpretations might be tested out on individuals and groups.
- Unravelling the assumptions underlying informants' interpretations and testing claims against observations, secondary sources, evidence of decisions, and other views.
- Ensuring that informants know how the findings will be used and inviting them to critically review them before they are finalised.
- Using a range of methods, sources of information and theoretical perspectives in order to triangulate findings.

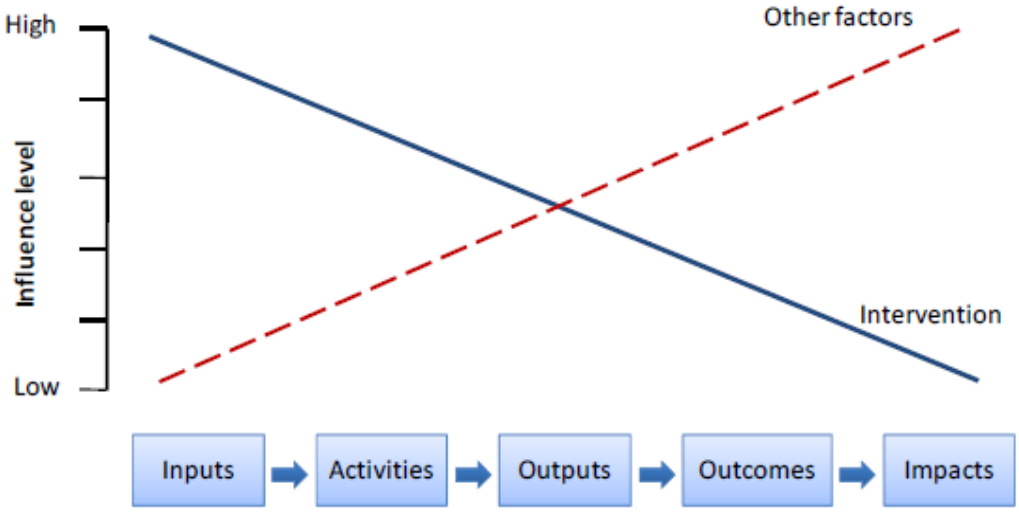
7.3 Attribution

Attribution is a central objective of evaluation of all types and particularly research impact evaluation. It refers to the ascription of a (causal) link between an intervention and an impact. An idea of the importance of attribution as an issue in research impact evaluation can be gained from the number of studies commenting on it. The following sources considered in this review make more than a passing reference to the issue: 151S, 152S, 153S, 154S, 156S, 161S, 163S, 166S, 169S, 170S, 181S, 185S, 197S, 201S, 202S, 203S, 204S, 206S, and 207S.

In discussing attribution there is a danger that it becomes a catch-all for the issues associated with evaluation, for example, how impacts are measured, whether they can be measured, and the impact on evaluation of the nature and relationship of the knowledge and policy systems and so on. Here we cover the main points raised in the literature in regard to attribution

The ability to quantify and establish attribution of impacts declines with the number of steps from the intervention - as an increasing number of other factors can be involved in the creating the impact. This is indicated graphically in Figure 14. It also needs to be remembered that the process of interventions leading to impacts takes time, sometimes a period of many years, and this is not explicit in Figure 14.

Figure 14: Relative influence of an intervention and other factors in the logic chain
(Adapted from Smutylo 2001. Original Source: Figure 45, 152S)



If an attribution of an impact to an intervention is to be confirmed, consideration needs to be given to the level of attribution desired. 152S suggests four levels of attribution in health research: attribution to research activities generally; research in a single discipline; research from a particular country and finally individual projects and programmes. It argues that most difficulties arise in trying to attribute outcomes in the last of these levels.

Methods of attribution

The methods described above can be used to collect the data elements that allow attribution to be explored or made, for example, information on research interventions, the impacts that could have resulted and the pathways by which these may be linked. The attribution itself can be based on a number of approaches. The literature suggests these include qualitative assessment, i.e. asking people how strongly outcomes are related to interventions (151S), econometrics or other statistical methods to demonstrate a link between the intervention and impact (152S, 174S), or pattern matching approaches (202S). 196S has an interesting section on the links between outputs and different types of outcomes and the methods used to demonstrate linkages between for 16 studies. It concludes that:

> ‘as the standard method progresses from inputs to outputs to outcomes in the literature, the results become less defined and the methods to measure them become more abstract…links to intermediate outcomes, such as public policy, regulations, interventions, and clinical guidelines, have been posited and are of considerable interest but actual studies focus on a limited set of specific cases or subjects’ (Section 6.0, 196S)

High quality data are required particularly for quantitative approaches. The potential value of logic models in exploring attribution is highlighted (152S and 202S). 153S uses a

combination of hypothesised impact pathways, key-informant interviews to verify the model, and cross checking both with documentary sources and between interviewees to explore attribution mechanisms.

It would therefore seem from an inspection of the literature that much attribution is actually based on establishing plausible links between interventions and impacts, for example, by statistical approaches or logic models. While links may be demonstrated, the causal mechanisms by which impacts occur are less clear. So it seems that while the term attribution implies a causal basis many evaluations are really only ascribing or linking impacts to interventions. As 152S suggests, establishing that an intervention has 'contributed' to an impact may therefore be a more appropriate term. However, as stated in 181S,

> 'While causality may be impossible to prove, it may often be inferred with some confidence.'

This is perhaps inevitable given the broader nature of the policy system, the complex relationship between research, policy and knowledge, and the subjective and conflicting accounts of impact and their causes.

166S comments on a further problem of attribution; research and policy are not independent of one another. Policy-makers can commission research which they specifically want and are therefore more likely to take up. It also considers other biases that can be introduced or overlooked in research impact evaluation analysis such as omitted variable bias, i.e. not controlling for others variables that may affect outcomes and selection bias, i.e. utilising results that support a hypothesis and disregarding others, and gives examples of these.

169S and 197S highlight the problem of measuring the impact of 'confirmatory research'. How is impact attributed to research that confirms the existing basis of a policy? 202S highlights that impact cannot be attributed to factors not included in a logic model, which can result in possible over-attribution of impact to the factors that are included. The problems of under- and over-attribution are also raised in 161S, 163S.

174S raises four methodological issues associated with existing conceptual and empirical studies on the use of research in government agencies:

1. Composition of the study population (needs to sample multiple policy domains, multiple government agencies of different sizes, and multiple levels of responsibility);

2. Specification of the dependent variable 'use' (there is too much focus solely on 'instrumental' use, need to consider other scales of use);

3. Problems associated with the independent variables considered, which tend to look more like checklists of variables assumed to explain utilisation (e.g. characteristics of

research products, dissemination efforts and links and exchanges between researchers and users etc), rather than formal heuristic devices .

4. Problems resulting from failure to appreciate respondents' inability to report and explain their behaviour accurately.

7.4 Research, knowledge and policy

As noted in the Defra review, the relationship between knowledge and decision making has been a matter of interest for hundreds, if not thousands, of years (Shadish *et al*., 1991). The diverse models of the relationship between policy and research have consequences for the evaluation of research impact. An example of this can be seen in Case Study 1, which outlines the Stockholm Environment Institute's approach to research impact evaluation (see the accompanying Case Studies report).

Building on other authors' contributions, Garrett and Islam (164S) capture succinctly the complexity of the interfaces between research, knowledge and policy and the issues raised by this for research impact evaluation. They argue against the view that research only has an impact on policy if there is a clear, direct link between research and policy outcomes. They suggest that, whilst the most striking examples of impact result when a policy-maker directly takes up research recommendations, this does not happen frequently, thus the impact of research on policy choice is more likely not to be direct or linear.

They suggest (as others do, e.g. 161S, 203S) that a research organisation's greatest influence probably occurs by contributing high-quality policy-relevant information to a pool of knowledge that policy-makers access when they need it and use as they see fit. i.e. measuring 'impact' of research by looking only at visible policy choices or policy outcomes would be misleading and should not be the principle metric for judging the impact of a research institute.

Research is one of many competing sources of information, which is itself one of many factors that affect the final policy decision. Policy-makers thus frequently use research less to dictate specific solutions than to help them think about issues and define the scope of problems and possible responses i.e. research brings insight into the policy process. Much of this research use is not deliberate or direct and does not correspond to specific pieces of research; rather, bits of information have seeped into the mind, uncatalogued, without citation. Information serves, cumulatively over time, a diffuse 'enlightenment' function, providing an understanding and interpretation of the data and the situation that is critical to the policy decision. As ideas from research become absorbed into conventional wisdom, they shape people's assumptions about how things work, about what needs to be done, and what solutions are likely to achieve desired ends.

We note the importance of this issue here which was discussed in more detail in a working paper produced during the project. This paper, produced by project partner Matthijs Hisschemöller at IVM, does not seek to make a decisive claim with respect to the state of the art in knowledge for policy research but rather aims to frame the complexity of the research-policy interface in such a way that it can help SKEP members to find their way in assessing the usefulness of different frameworks and methods. Hisschemöller (2009) points out that for several decades, policy scientists and sociologists have been struggling with the conceptualisation of knowledge use in public policy. There have been extensive efforts to structure this issue in a manageable way, which have resulted in quite different frameworks for evaluation research. Notwithstanding all differences, he concludes that there is scientific consensus with respect to at least two things.

Firstly, the variables that need to be taken into account in research evaluation relate to:

- The type of knowledge, which may also include institutional factors such as culture and tradition with respect to academic disciplines and the institutionalisation of academic research.
- The institutional policy context, including policy process, characteristics of the policy subsystem including the interactions between knowledge providers and potential users.
- Different types of actual research use.

Secondly, there are a broad range of variables making it difficult to actually measure the uptake of knowledge in an objectified manner.

8 Are there any specific issues for evaluating the impact of research on environmental policy regulatory uptake and implementation?

A large amount of the research impact evaluation literature is based on theory and practice from the health sector. The question therefore arises whether this practice and theory is transferable to environmental research and policy.

The previous Defra review identified two issues of transferability. 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 is that environmental research can be a politically charged area as it often implies the need for changes in attitude and behaviour. In some policy areas, evaluators need to be particularly 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. Some areas are clearly more mature in policy terms and evaluators might therefore expect to identify greater policy activity.

Environment and sustainable development policy is a particularly challenging area of policy development, given its rate of development, 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).

A further issue of transferability of the HERG model is highlighted from 208S. Their conclusion on the transferability of the HERG payback approach is that it 'is more easily applicable where the processes of research, application and impact follow a linear progression and are carefully documented'.

The nature of environmental research means this is often unlikely to be the case. For example, 167S highlights issues of uncertainty in environmental science and the problems this causes for policy-makers; the multi-disciplinary nature of much environmental research; and the legacy of past prediction errors could cause problems in adaption of the HERG approach to environmental policy. 188S also highlights the importance of scale and location in environmental research and policy response to it.

194S captures the problems associated with the use of environmental research and its analysis, which obviously poses challenges for how its impacts may be evaluated. It highlights that environmental challenges may be complex, trans-scientific and 'unstructured'

and considers the implications of this for its use. Different kinds of knowledge are deemed to be relevant (expert and lay, universal and contextual, technical and cultural), and new framings of problems may allow previously rejected solutions to be attached to them. Within this framework, research is related to policy but not in a linear fashion; '*it creeps or is absorbed into…policy…via indirect, cumulative and diffuse processes…and in combination with lay knowledge'*. Often knowledge will appear to lie dormant until external events facilitate policy change. Such processes tend to be long-term ones covering periods of a decade or more, and invariably involve learning of various kinds.

The work of the Royal Commission on Environmental Pollution on integrated pollution control and public access to environmental information is also noted in 194S. The research initially fell on unpromising ground but ultimately came to be reflected in lasting policy change.

In spite of searching for it the literature synthesised in this review does not reveal much information about the impact of research on environmental regulation and uptake specifically as opposed to policy impact generally. 196S is a recent review which aims to assess research impact assessment or evaluation of grant programmes, particularly in the area of health or environmental health. It indicates that though the development of regulations was noted in the literature as a possible immediate outcome of research, none of the studies reviewed proposed methodologies for linking regulations to research. Instead, they simply linked regulations to downstream outcomes (e.g. economic cost/benefit for the implementation of regulation). However, the paper suggests that methodologies similar to that used for policy analysis (i.e. interviews with decision-makers) may also be applicable to regulations. This seems likely to be a reasonable assumption.

The insights from the literature on the relationship between knowledge and impact, whether policy or implementation would also seem to be relevant. What is the typical relationship interaction between regulation and knowledge generated by research. For example, are regulations typically derived from research with clear specific objectives and standards to be set and therefore more instrumental in nature, and/or is a interpretation stage required in which conceptual insights into regulatory standards and requirements are translated into standards and uptake?

9 What conceptual and/or theoretical studies underpin research impact evaluation?

9.1 The importance of conceptual and theoretical models

As emphasised in both this review and the Defra review, adopting and/or adapting a theoretical model of the relationship between research and policy is considered to be a valuable step in the development of a plan for studying the impact of research upon policy. A number of the papers in this review highlight the value to individual research evaluations of starting with a common, conceptual framework that is compatible with existing frameworks/approaches (79a, 152S, 170S, 181S, 209S, and 203S).

Broadly speaking there are two types of theoretical model relevant to research impact evaluation. The first type tends to visualise the relationship between research and policy and the processes by which research may impact on policy. The second type includes frameworks designed to facilitate evaluation of research impacts which focus more on the process and needs of evaluation methods.

This section of the report considers the theoretical models that underpin the evaluation of the impact of research studies. It draws on three sources: the previous review (which had identified a number of theoretical models), an analysis of the theoretical and conceptual approaches discussed and used in the new papers identified through for this review, and the working paper prepared by a member of the team with considerable expertise in developing and using theoretical models of the relationship between research, knowledge systems and policy. Given the nature of this area the theoretical models discussed are largely derived from empirical observation.

The Defra review

The previous review for Defra highlighted three models in particular that were specifically developed for the purpose of evaluating the impact of research studies. These were the Health Economics Research Group (HERG) Payback Model, the RAPID Outcome Assessment and the Research Impact Framework (RIF). Details of these are given above.

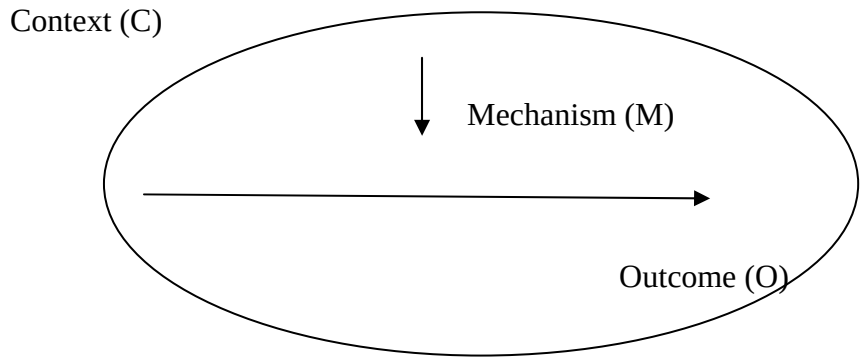
These approaches suggest two key dimensions that need to be present in a model of research impact evaluation:

- An interest in capturing **Context** (for example, in HERG through case studies and in RIF through narrative accounts.)
- The need to evaluate **Outcomes** – i.e. going beyond research outputs

These dimensions immediately share some ground with Realist Evaluation which applies to evaluation more broadly, rather than the narrower assessment of the impact of research on

(environmental) policy and uptake. As illustrated in Figure 15, Pawson and Tilley (1997)[10] encapsulate their approach in what they call the CMO configuration: it is only possible to understand outcomes (O) by taking into consideration the mechanisms (M) at work and the context in which the intervention or programme is tried and tested (C).

Figure 15: The Context Mechanism Outcome (CMO) Configuration



Pawson and Tilley contribute an understanding of **Mechanisms** to the discussions of context and outcomes in the previous models of research impact evaluation. Mechanisms are what make change happen in a particular system and as the review for this project has found there are numerous direct, and indirect, mechanisms that cause impact.

The value of this model lies in its focus on outcomes and the generation of understanding of causality in complex systems, through an exploration of context and change mechanisms. However, the CMO configuration is a generic evaluation model. In the context of research impact, the existing models discussed in the previous review urge us to also consider:

- The need to explore **Links –** between different stakeholders in the process,
- The need to describe **Process** (the pathways or pipelines by which research achieves impacts),
- The need to consider the **Quality** of the research, or what Crewe and Young (2002) (158S) describe under Evidence as ‘approach and credibility’.

The SKEP review

The papers identified in this updated and expanded review also discussed the HERG, RAPID and RIF models, but added a number of new approaches. In particular, the Australian Research Quality Framework (RQF) and the Irish Environmental Protection Agency approach. They also discuss the economics-based approaches used by CGIAR and IFPRI. The new review identified a further set of papers that discuss theories of how research is used in policy (largely where this theory has been used to inform the adopted approach to research impact evaluation).

[10] Pawson R and Tilley N (1997) Realistic Evaluation, Sage, London

By far the most influential and widely referenced work in this area is by Carol Weiss[11], who published a paper in 1979 identifying different approaches to research use (from the political and tactical to the enlightened). Recent work by Nutley and colleagues (108) (in particular their continuum of research use) is referenced in a number of papers. The working paper for this project by Matthijs Hisschemöller discusses some of the key theoretical contributions. Four themes have been drawn from the theoretical papers visualising the relationship between research and policy (rather than building models themselves). These are discussed below and are: critical factors affecting impact; depth of impact; types of research use; and, pathways of uptake.

9.1.1 Critical factors affecting impact.

A number of studies begin to group the variables cited in the literature as determining research impact/uptake. For example 174S suggest the following categories of explanations of research uptake: engineering explanations; socio-organisational explanations; the 'two communities' idea (which assumes that a difference between the culture of professionals and managers and the culture of university researchers leads to a lack of communication between them); the 'interaction' explanation (suggests that a lack of interaction between researchers and their potential audiences leads to under-utilisation of research findings); and individual attributes, such as respondents' level of responsibility and level of education. This links to the Mechanisms, Context and Links aspects of the Pawson and Tilley model.

9.1.2 Depth of impact

Depth of impact is also discussed in a number of studies (see for example 174S). This study touches on the idea of 'scales of knowledge utilisation' – in this study they use a cumulative six-stage scale including reception, cognition, discussion, reference, effort and influence. This links to **Outcomes** aspect of the model.

9.1.3 Types of research use

The utilisation of research by policy-makers is often framed in terms of engineering (the direct application of research results to a pending decision) versus enlightenment (the indirect influence of research on the ways in which we think about social problems and how they might be tackled). There seems to be general support for the view that in the policy arena, research is most likely to have an enlightenment rather than engineering impact (189 S, 194S, 195S, 201S, 205S). Nutley et al (2007) identify a continuum of impact from conceptual to instrumental, ranging from raising awareness, through shifts in knowledge, attitudes and culture, to actual changes in day-to-day practice. This continuum is seen as a two-way process, recognising that the use of evidence is likely to be iterative and interactive rather than necessarily straightforwardly linear. When discussing the 'impact assessment' model 190S

[11] Weiss, Carol H. 1979. *The Many Meanings of Research Utilization, Public Administration Review*, 39 (5): 426-431.

draws on Nutley *et al.*'s (108) concept of a continuum, from conceptual to instrumental use of research in policy, and suggests that strategies to enhance research impact may address any point on this continuum. This links to the **Outcomes** aspect of the model.

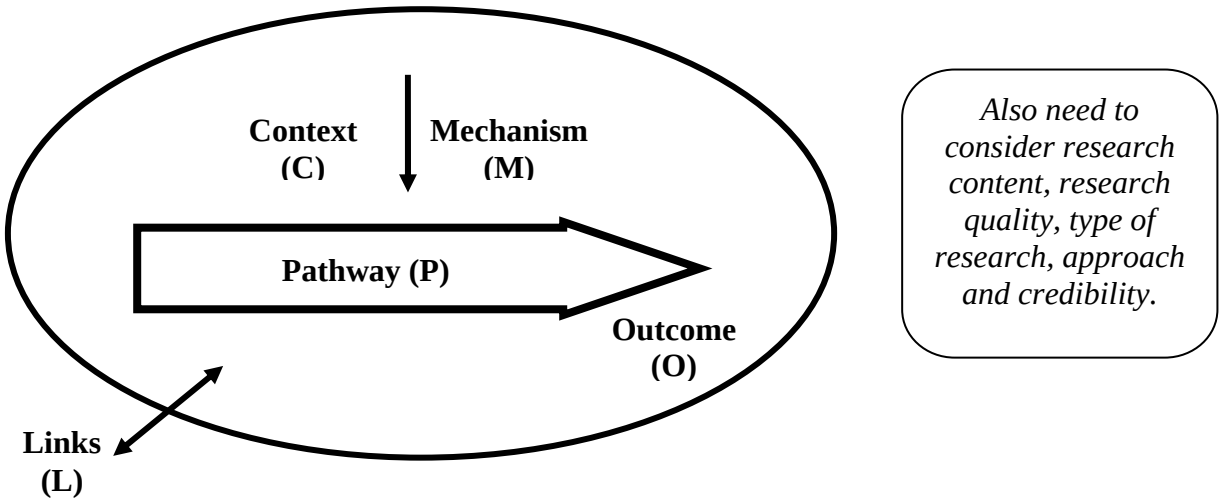
9.1.4 Pathways of uptake and logic models

The value of using logic models in evaluations is highlighted in 154S. The tendency for policies and programmes to place little attention on expectations with respect to outcomes and impacts, i.e. big chunks of the logic model are often missing or poorly defined, thereby hindering subsequent evaluation attempts, is also highlighted. 201S suggests the influence and impact of research can be generally understood through the concept of 'impact pathways' – each pathway describes a set of causal relationships that link research to the eventual impact. Research innovations may progress down multiple pathways simultaneously. Within any given impact pathway, a multitude of factors determine whether or not the changes in behaviour, represented as a series of causal links, will occur. Research 'influence' may occur that never leads to impact, for example, science findings are used in public debate but that debate does not lead to a policy reform, or research leads to a policy reform but regulatory agencies fail to implement the change. This links to the **Pathway** aspect of the model.

9.2 Key dimensions of a conceptual model for an applied model of impact evaluation

This consideration of the theoretical underpinning of research impact evaluation points to a development of the CMO model to explicitly acknowledge links and pathways – 'CMO+2' (Context, Mechanisms, Outcome, Links and Pathways) as suggested below in Figure 16. These and the more detailed comments above need to be reflected in any practical approach to research impact evaluation and the development of the Guide and Guidelines.

Figure 16: The CMO+2 Conceptual Model of Evaluation



10 Conclusions from the literature review

This review has revealed a significant amount of new material and detail on evaluating the impact of research on policy which was not identified in the previous Defra review. Reassuringly, the new material is consistent with the previous review. The main findings of the Defra review in relation to the question 'how do people evaluate the impact of research?' included:

1. Evaluation methods used should be 'fit for purpose', yet the variety of methods and uses for them means this decision is far from straight forward. Approaches are therefore context dependent with an important factor being the outcome of interest.

2. Much of the research impact evaluation that has been done has focused on non-policy impacts, i.e. on different outcomes of interest to the one we are interested in for this project, particularly in the area of health and clinical practice of healthcare options. This difference needs to be considered in developing policy-focused evaluations.

3. Context also affects the type of evaluation conducted. For example, in the area of international development the emphasis has been on qualitative, participatory evaluations intended to improve learning and service delivery with frequent field visits to establish practice on the ground. In other contexts, for example the United States, the emphasis has been on developing quantitative methods and performance indicators for statutory reporting requirements. In the European Union panel review has been the main method of evaluation and, while criticised in some regards, its role in building ownership across countries is important.

4. There is a common interest in developing new approaches to evaluation, for example, through network analysis and citation analysis. However, the differing contexts and needs of users of these methods are likely to result in an ongoing diversity of approaches, and new methods being used that complement rather than replace existing methods.

The first finding of the Defra review identifies a problem which this review has attempted to address through the collation of material relevant to the planning and development of a research impact evaluation. The previous review identified a wide range of options and highlighted the difficulty of choosing between them. This review has brought together material on the issues to consider and address in planning a review in detail. These include the questions of why, what, when and who to evaluate, which are explored in Section 3.1, and key practical issues to be addressed in developing an evaluation framework (Section 4) which can be used to test any evaluation approaches developed.

In relation to the second finding of the Defra review, about most documented evaluations being on non-policy impact, this review has identified evaluation methods successfully

applied to the evaluation of policy impacts generally and environmental policy specifically. These include, for example, the Irish EPA approach and Research Impact Framework. The Case Studies also give a range of examples of evaluation approaches use to understand the impact of research on environmental policy including examples from the Environment Agency (England and Wales), the Swedish Environmental Protection Agency and the Finnish Environment Institute (SYKE).

This review gives greater detail of the approaches identified than in the previous review and has brought together detailed information on five evaluation frameworks thought to be of particular relevance to this project's objectives. It has identified some additional data collection methods and approaches to analysis. It also adds new insights which are important to the project's aim of developing an approach for the evaluation of the implementation and uptake of environmental research and practical guidance for its use.

11 References

The individual sources are referred to by a unique reference number. We have continued the numbering of references from the Defra review into this Review to avoid the potential confusion of numbers being duplicated.

We have suffixed the numbers of all sources identified during this review with an 'S' (for SKEP), for example 149S. The number of the first reference identified for this review is 147S, not 157S, as might be expected given the 156 references reviewed in the Defra review. This is because for cataloguing reasons some numbers in the Defra review were suffixed by letters, e.g. 26a, 26b which meant the last number used in the Defra review was 146.

1. Acosta A and Douthwaite B (2005) *Appreciative Inquiry: an approach for learning and change based on our own best practices*, ILAC Brief 6, Rome: Institutional Learning and Change. http://www.cgiar-ilac.org/downloads/Briefs/Brief6Proof2.pdf.

2. Adamo A (2002) *Strategic evaluation of policy influence: what evaluation reports tell us about public policy influence by IDRC-supported research*. Ottawa: International Development Research Centre. http://www.idrc.ca/uploads/user-S/10359910880strategic.pdf.

3. Adamo A (2003) *Influencing public policy through IDRC-supported research: synthesis of document reviews*. Ottawa, International Development Centre, Evaluation Unit. http://www.idrc.org.sg/en/ev-31926-201-1-DO_TOPIC.html.

4. Allen P, Peckham S, Anderson S and Goodwin N (2007) Commissioning research that is used: the experience of the NHS Service Delivery and Organisation Research and Development Programme. *Evidence & Policy* 3(1): 119-134.

5. Almeida C and Bascálo E (2006) Use of research results in policy decision-making, formulation, and implementation: a review of the literature. *Cadernos de Saúde Pú blica* 22(Suppl): S7-S19. http://www.scielosp.org/pdf/csp/v22s0/02.pdf.

6. Anderson R (2006) The policy impact of population health surveys: an illustration of the measurement challenges using the NSW Health Survey. *Evidence & Policy* 2(2): 167-183.

7. Annema J A, van Wee B, van Hoek T and van der Waard J (1999) Evaluation of Dutch public investment plans. *Environmental Impact Assessment Review* 19(3): 305-317.

8. Arnold E (2004) Evaluating research and innovation policy: a systems world needs systems evaluations. *Research Evaluation* 13(1): 3-17.

9. Arnold E, Muscio A and Zaman R (2005) *Reviews reviewed: lessons from the first six international panel reviews 1999-2004,*. Brighton: Technopolis, for the Engineering and Physical Sciences Research Council. http://www.epsrc.ac.uk/CMSWeb/Downloads/Publications/Other/InternationalReviewEvaluation.pdf.

10. Arts and Humanities Research Council (2007) *AHRC impact strategy*. Bristol: Arts and Humanities Research Council. http://www.ahrc.ac.uk/images/AHRC_impact_strategy.pdf.

11. Beacham B, Kalucy L and McIntyre E (2005) Understanding & measuring research impact. *FOCUS on...* (2) 12pp. http://www.phcris.org.au/publications/focuson/index.php.

12. Bekker M P M, Putters K and van der Grinten T E D (2004) Exploring the relation between evidence and decision-making: a political-administrative approach to health impact assessment. *Environmental Impact Assessment Review* 24(2): 139-149.

13. Berryman J (2006) What defines 'enough' information? How policy workers make judgements and decisions during information seeking: preliminary results from an exploratory study. *Information Research* 11(4) unpaged. http://informationr.net/ir/11-4/paper266.html.

14. Big Lottery Fund (2005) *Increasing the impact of research, evaluation and learning.* London: Big Lottery Fund. http://www.biglotteryfund.org.uk/er_eval_increase_impact_report.pdf.

15. Biotechnology and Biological Sciences Research Council (2005) *BBSRC strategy for evaluating research programmes*, SB 44/2005. Swindon: BBSRC. http://www.bbsrc.ac.uk/organisation/structures/boards/strategy/2005/0511_evaluating_research.pdf.

16. Boaz A (2006) *Systematic reviews as a source of evidence for policy: an in-depth exploration of a review on mentoring*. PhD Thesis. London.

17. Bouma J and Droogers P (2007) Translating soil science into environmental policy: a case study on implementing the EU soil protection strategy in the Netherlands. *Environmental Science and Policy* 10(5): 454-463.

18. Bozeman B (2003) *Public value mapping of science outcomes: theory and method. A monograph of the Public Value Mapping Project of the Center for Science, Policy and*

Outcomes. Washington, DC: Columbia University, Center for Science, Policy and Outcomes, and Atlanta: Georgia Tech, School of Public Policy. http://www.cspo.org/home/cspoideas/know_flows/Rock-Vol2-1.PDF.

19. Brehaut J D and Juzwishin D (2005) *Bridging the gap: the use of research evidence in policy development*, HTA Initiative 18. Edmonton: Alberta Heritage Foundation for Medical Research. http://www.ihe.ca/documents/hta/HTA-FR18.pdf.

20. Brofoss K E (1998) The Research Council of Norway's use of research evaluation: an assessment of research evaluation as a strategic tool. *Research Evaluation* 7(3): 134-140.

21. Bumgarner R, Hattaway D, McGann J and Wise H (2006) *Center for Global Development: evaluation of impact*. Washington, DC: Arabella Philanthropic Investment Advisors. http://www.arabellaadvisors.com/CGDreport/CGDWeb.pdf.

22. Butler L and McAllister I (2007) *Metrics or peer review? Evaluating the 2001 UK Research Assessment Exercise in political science*. Canberra: Australian National University, Research School of Social Sciences. http://repp.anu.edu.au/papers/2007_ukresearchassess.pdf.

23. Buxton M, Hanney S, Packwood T, Roberts S and Youll P (2000) Assessing benefits from Department of Health and National Health Service research & development. *Public Money & Management* 20(4): 29-34.

24. Campbell S, Benita S, Coates E, Davies P and Penn G (2007) *Analysis for policy: evidence-based policy in practice*. London: Government Social Research Unit. http://www.gsr.gov.uk/downloads/resources/pu256_160407.pdf.

25. Carden F (2004) Issues in assessing the policy influence of research. *International Social Science Journal* 56(179): 135-151.

26. Carden F (2005) *Making the most of research: The influence of IDRC-supported research on policy processes*, Paper for the International conference 'African Economic Research Institutions and Policy Development: Opportunities and Challenges', Dakar, January 28-29 2005. http://www.idrc.ca/uploads/user-S/11085518871Making_the_Most_of_Research.pdf.

26a. Carden, F (2007) Context Matters – the influence of IDRC-supported research on policy processes, in Ayuk, E T and Marouani, M A (2007) *The policy Paradox in Africa: Strengthening links between economic research and policymaking*. New Jersey: Africa World Press and Ottawa: International Development Research Centre:

26b. Carden, F (Forthcoming, 2008) *Using comparative data: A systems approach to a multiple case study*, chapter for a book on case methods

26c. Carden, F (2008) The Episode Studies Approach. in Trussie, D *The Politics of Trade: the role of research in trade policy and negotiation.* Amsterdam and Ottawa: Martin Nijhoff-Brill and International Development Research Centre.

26d. Carden, F (Forthcoming, 2008) *Knowledge to Policy:Making the most of development research*. Ottawa: Sage & International Development Research Centre (IDRC)

27. Cash D and Clark W (2001) *From science to policy: assessing the assessment process*, Faculty Research Working Paper RWP01-045. Cambridge, MA: Harvard University, John F Kennedy School of Government. http://belfercenter.ksg.harvard.edu/publication/by_type/working_paper.html.

28. CGIAR Science Council (2006) *Impact assessment of policy-oriented research in the CGIAR: a scoping study report*. Rome: CGIAR Science Council Secretariat. http://www.sciencecouncil.cgiar.org/publications/pdf/CG_SC_PORIA-pr2(enabl).pdf.

29. Council for the Humanities, Arts and Social Sciences (2005) *Measures of quality and impact of publicly funded research in the humanities, arts and social sciences*, CHASS Occasional Paper 2. Canberra: Council for the Humanities, Arts and Social Sciences. http://www.chass.org.au/papers/pdf/PAP20051101JP.pdf.

30. Christiansen K with Hovland I (2003) *The PRSP initiative: multilateral policy change and the role of research*, Working Paper 216. London, Overseas Development Institute. http://www.odi.org.uk/publications/working_papers/wp216.pdf.

31. Church C (2005) *Mind the gap: policy development and research on conflict issues*. Londonderry: University of Ulster, International Centre of Excellence for Conflict and Peace Studies. http://www.incore.ulst.ac.uk/policy/rip/RIP.pdf.

32. Clark W, Mitchell R, Cash D and Alcock F (2002) *Information as Influence: how institutions mediate the impact of scientific assessments on global environmental affairs*, Faculty Research Working Paper RWP02-044. Cambridge, MA: Harvard University, John F Kennedy School of Government. http://ksgnotes1.harvard.edu/research/wpaper.nsf/rwp/RWP02-044/$File/rwp02_044_clark.pdf.

33. Committee on Science, Engineering, and Public Policy (2001) *Implementing the Government Performance and Results Act for research. a status report*. Washington, DC: National Academy Press.

http://www.nap.edu/catalog.php?record_id=10106.

34. Council for Medical Sciences (2002) *The societal impact of applied health research: towards a quality assessment system*. Amsterdam: Royal Netherlands Academy for Arts and Sciences. http://www.knaw.nl/publicaties/pdf/20021098.pdf.

35. Cozzens S (1995) US research assessment: recent developments. *Scientometrics* 34(3): 351-362.

36. Cranfield S and Ward H (2002) *'Managing Change in the NHS' (2001) publications: report of a study on the uptake and use of two SDO publications*. London: National Co-ordinating Centre for the NIHR Service Delivery and Organisation Programme (NCCSDO). http://www.sdo.lshtm.ac.uk/files/adhoc/change-management-survey.pdf.

37. Cunningham P, Boden M, Glynn S and Hills P (2001) *Measuring and ensuring excellence in government science and technology: international practices. France, Germany, Sweden and the Unired Kingdom*. Manchester: University of Manchester PREST (Policy Research in Engineering, Science and Technology). http://www.mbs.ac.uk/Research/engineeringpolicy/publications/documents/CSTA-PREST.pdf.

38. Davies H, Nutley S and Walter I (2005) *Approaches to assessing the non-academic impact of social science research: report of the ESRC symposium on assessing the non-academic impact of research*, St Andrews: University of St Andrews, Research Unit for Research Utilisation. http://www.odi.org.uk/rapid/Meetings/ESRC/Index.html.

39. Davies R and Dart J (2005) *The 'Most Significant Change' (MSC) technique: a guide to its use*. London: CARE International et al. http://www.mande.co.uk/docs/MSCGuide.pdf.

40. Deelstra Y, Nooteboom S G, Kohlmann H R, van den Berg J and Innanen S (2003) Using knowledge for decision-making purposes in the context of large projects in the Netherlands. *Environmental Impact Assessment Review* 23(5): 517-541.

41. Delanghe H and Muldur U (2007) Ex-ante impact assessment of research programmes: the experience of the European Union's 7th Framework Programme. *Science and Public Policy* 34(3): 169-183.

42. Earl S and Carden F (2002) Learning from complexity: the International Development Research Centre's experience with Outcome Mapping. *Development in Practice* 12(3/4): 518-524.

43. Earl S, Carden F, Smutylo T (2001) *Outcome Mapping: building learning and reflection into development programs*. Ottawa: International Development Research Centre. http://www.idrc.ca/en/ev-9330-201-1-DO_TOPIC.html.

44. Edwards K (2001) *PCRs and policy influence. What project completion reports have to say about public policy influence by Centre-supported research: final report*. Ottawa: International Development Research Centre, Evaluation Unit. http://www.idrc.ca/uploads/user-S/10359905080pcr_short_final.pdf.

45. Elliott H and Popay J (2000) How are policy makers using evidence? Models of research utilisation and local NHS policy making. *Journal of Epidemiology and Community Health* 54(6): 461-468.
46. European Commission (2005) *Assessing the impact of energy research*, EUR 21354. Luxembourg: Office for Official Publications of the European Communities. http://ec.europa.eu/research/energy/pdf/erevia_en.pdf.

47. Evans J P (2006) Lost in translation? Exploring the interface between local environmental research and policymaking. *Environment and Planning A* 38(3): 517-532.

48. Fahrenkrog G, Polt W, Rojo J, Tübke A and Zinöcker K (eds) (2002) *RTD Evaluation Toolbox: assessing the socio-economic impact of RTD-policies*, IPTS Technical Report Series, EUR 20382. Seville: European Commission, Joint Research Centre, Institute for Prospective Technological Studies. http://ftp.jrc.es/eur20382en.pdf.

49. Fisher E and Holland J D (2003) Social development as knowledge building: research as a sphere of policy influence. *Journal of International Development* 15(7): 911-924.

50. Flint M E S (1998) *Natural resources research evaluation synthesis*, Evaluation Report EV:604. London: Department for International Development. http://www.dfid.gov.uk/aboutdfid/performance/files/ev604.pdf.

51. Gardner B L (1999) *Returns to policy-related social science research in agriculture*, Impact Assessment Discussion Paper No. 9. Washington, DC: International Food Policy Research Institute. http://www.ifpri.org/impact/iadp09.pdf.

52. Garfinkel M S, Sarewitz D, Porter A L (2006) A societal outcomes map for health research and policy. *American Journal of Public Health* 96(3): 441-446.

53. General Accounting Office (2003) *Pipeline safety: systematic process needed to evaluate outcomes of research and development program*, GAO-03-746. Washington, DC: United States General Accounting Office. http://www.gao.gov/new.items/d03746.pdf.

54. Government Accountability Office (2007) *US public diplomacy: actions needed to improve strategic use and coordination of research*, GAO-07-904. Washington, DC: United States Government Accountability Office. http://www.gao.gov/new.items/d07904.pdf.

55. Georghiou L (1995) Assessing the Framework Programmes: a meta-evaluation.

Evaluation 1(2): 171-188.

56. Georghiou L and Davis E (eds) (1988) *Evaluation of R & D: a policymaker's perspective: proceedings of the International Workshop on Assessment and Evaluation, Queen Elizabeth II Conference Centre London*, London: HMSO.

57. Gerhardus A, Dujardin M, Kiet P H T, Siddiqi S and Sauerborn R (2000) *A methodology to assess the use of research for health policy development*, Discussion Paper 04. Heidelberg: University of Heidelberg Medical School, Department of Tropical Hygiene and Public Health. For availablility, contact: agerhardus@yahoo.de.

58. Gerhardus A and Kielmann K (2000) The use of research for decision-making in the health sector: the case of 'Shared Care' in Burkina Faso, in: *Lessons in research to action: case studies from seven countries*. Geneva: Council on Health Research for Development: 19-27. http://www.cohred.org/main/CommonCategories/content/696.pdf.

59. Godin B and Doré C (2003) *Measuring the impacts of science: beyond the economic dimension*. Montreal: Canadian Science and Innovation Indicators Consortium. http://www.csiic.ca/PDF/Godin_Dore_Impacts.pdf.

60. Government of Western Australia (1997) Appendix B: PIs for policy development and advice, in: *Preparing performance indicators: a pratical guide*. Perth: Government of Western Australia, Department of the Premier and Cabinet, Public Sector Management Office. http://www.dpc.wa.gov.au/psmd/pubs/psrd/preppis.html.

61. Grosse S D, Teutsch S M and Haddix A C (2007) Lessons from cost-effectiveness research for United States public health policy. *Annual Review of Public Health* 28: 365-391.

62. Guy K, Amanatidou E and Psarra F (2005) Framework Programme 5 (FP5) impact assessment: a survey conducted as part of the five-year assessment of European Union research activities (1999-2003). *Science and Public Policy* 32(5): 349-366.

63. Hanney S, Davies A and Buxton M (1999) Assessing benefits from health research projects: can we use questionnaires instead of case studies? *Research Evaluation* 8(3) pp189-199

64. Hanney S R, Gonzalez-Block M A, Buxton M J and Kogan M (2003) The utilisation of health research in policy-making: concepts, examples and methods of assessment. *Health Research Policy and Systems* 1(2). http://www.health-policy-systems.com/content/1/1/2.

65. Hanney S R, Grant J, Wooding S and Buxton M J (2004) Proposed methods for reviewing the outcomes of health research: the impact of funding by the UK's 'Arthritis Research Campaign'. *Health Research Policy and Systems* 2(4). http://health-policy-systems.com/content/2/1/4.

65a. Hanney S, Buxton M, Green C, Coulson D and Raftery J (2007) An assessment of the impact of NHS Health Technology Assessment Programme. *NHS Health Technology Assessment Programme*. Vol 11, number 53. http://www.ncchta.org/fullmono/mon1153.pdf

66. Healy R G and Ascher W (1995) Knowledge in the policy process: incorporating new environmental information in natural resources policy making. *Policy Sciences* 28(1): 1-19.

67. Herrick, C and Sarewitz D (2000) Ex post evaluation: a more effective role for scientific assessments in environmental policy. *Science, Technology, & Human Values* 25(3): 309-331.

68. Hilderbrand M and Simon J (2000) The role of research in child health policy and programs in Pakistan, in: *Lessons in research to action: case studies from seven countries*. Geneva: Council on Health Research for Development: 77-86. http://www.cohred.org/main/CommonCategories/content/696.pdf.

69. Her Majesty's Treasury, Department for Education and Skills, Office of Science and Technology and Department of Trade and Industry (2002) *Cross-cutting review of science and research: final report.* London, HM Treasury. http://www.hm-treasury.gov.uk/media/E/7/science_crosscutter.pdf.

70. Holmes J (2005) *The use of science in policy and regulation: baseline review*. Swindon: Environmental Research Funders Forum. http://www.erff.org.uk/reports/events.asp.

70a. Holmes, J (2007) *Study of dissemination and implementation of research in SKEP member organizations.* Scientific Knowledge for Environmental Protection (SKEP) http://www.skep-era.net/site/82.asp

71. Hooton N., Lee-Smith D, Nasinyama G and Romney D (2006) *Learning lessons from the Kampala urban agriculture policy process*, Process and Partnership for Pro-Poor Policy Change Project report. International Livestock Research Institute, and London: Overseas Development Institute. http://www.pppppc.org/content/files/documents/Kampala%20report%20final%20WEB.pdf.

72. Hovland I (2007) *Making a difference: M&E of policy research*, Working Paper 281.

London, Overseas Development Institute.
http://www.odi.org.uk/publications/working_papers/wp281.pdf.

73. International Food Policy Research Institute (2002) *Impact evaluation: assessing the impact of policy-oriented social science research*, Summary report of an IFPRI symposium workshop. Washington, DC, IFPRI.
http://www.ifpri.org/pubs/ib/ib5.pdf.

74. Innvær S, Vist G, Trommald M and Oxman A (2002) Health policy-makers' perceptions of their use of evidence: a systematic review. *Journal of Health Services Research and Policy* 7(4): 239-244.

75. Jensen F S and Koch N E (2004) Twenty-five years of forest recreation research in Denmark and its influence on forest policy. *Scandinavian Journal of Forest Research* 19(Suppl. 4): 93-102.

76. Johnston R (1995) Research impact quantification. *Scientometrics* 34(3): 415-426.

77. Jones S A, Fischhoff B and Lach D (1999) Evaluating the science-policy interface for climate change research. *Climatic Change* 43(3): 581-599.

78. Kalpazidou-Schmidt E (2006) RTD evaluation and policy in the European Research Area. *Evidence & Policy* 2(2): 185-209.

79. Kilpatrick H E (1998) *Some useful methods for measuring the benefits of social science research*, Impact Assessment Discussion Paper No. 5. Washington, DC: International Food Policy Research Institute. http://www.ifpri.org/impact/iadp05.pdf.

79a. Kivimaa, P, Mela, H, and Furman, E (2007) *Approaches and practices in mid-term and ex-post evaluation of environmental research programmes*. Scientific Knowledge for Environmental Protection (SKEP) Work Package 3, 28th June 2007 http://www.skep-era.net/site/files/WP3%20final_eval.v1.2.pdf

80. Kostoff R N (1994) Assessing research impact: US government retrospective and quantitative approaches. *Science and Public Policy* 21(1): 13-22.

81. Kostoff R N (1995) Research requirements for research impact assessment. *Research Policy* 24(6): 869-882.

82. Kostoff R N (1995) Federal research impact assessment: axioms, approaches, applications. *Scientometrics* 34(2): 163-206.

83. Kuhlmann S (1995) German government department's experience of RT&D programme evaluation and methodology. *Scientometrics* 34(3): 461-471.

84. Kuruvilla S, Mays N, Pleasant A and Walt G (2006) Describing the impact of health research: a Research Impact Framework. *BMC Health Services Research* 6(134). http://www.biomedcentral.com/content/pdf/1472-6963-6-134.pdf.

84a. Kuruvilla S, Mays N and Walt G (2007) Describing the impact of health services and policy research, *Journal of Health Services Research and Policy*. Vol 12 (Suppl 1) 23-31

85. Lavis J N, Ross S E, Hurley J E et al (2002) Examining the role of health services research in public policy-making. *Milbank Quarterly* 80(1): 125-154.

86. Lavis J N, Ross S E, McLeod C and Gildiner A (2003) Measuring the impact of health research. *Journal of Health Services Research & Policy* 8(3): 165-170.

87. Leksmono C, Young J, Hooton N, Muriuki H and Romney D (2006) *Informal traders lock horns with the formal milk industry: the role of research in pro-poor dairy policy shift in Kenya*, Working Paper 266. London: Overseas Development Institute. http://www.odi.org.uk/publications/working_papers/wp266.pdf.

88. Lewis J M (2005) A network approach for researching partnerships in health. *Australia and New Zealand Health Policy* 2(22). http://www.anzhealthpolicy.com/content/2/1/22.

89. Lindquist E A (2001) *Discerning policy influence: framework for a strategic evaluation of IDRC-supported research*. Victoria, BC: University of Victoria, School of Public Administration. http://www.idrc.ca/uploads/user-S/10359907080discerning_policy.pdf.

90. Link A (1995) Economic performance measures for evaluating government-sponsored research. *Scientometrics* 36(3): 325-342.

91. Luukkonen T (1998) The difficulties in assessing the impact of EU Framework Programmes. *Research Policy* 27(6): 599-610.

92. Luukkonen-Gronow T (1987) Scientific research evaluation: a review of methods and various contexts of their application. *R&D Management* 17(3): 207-221.

93. Luukkonen-Gronow T (1989) The impact of evaluation data on policy determination, in: D Evered and S Harnett (eds) *The evaluation of scientific research*. London: Ciba Foundation: 234-246.

94. Mackinnon J and Reinikka R (2002) How research can assist policy: the case of economic reforms in Uganda. *World Bank Research Observer* 17(2): 267-292.

95. Marburger J H (2005) Wanted: better benchmarks. *Science* 308(5725): 1087.

96. McNie E C (2006) Reconciling the supply of scientific information with user demands: an analysis of the problem and review of the literature. *Environmental Science and Policy* 10(1): 17-38.

96a. Meagher L, Lyle C and Nutley S (submitted 2008) *Flows of knowledge, expertise and influence: a method for assessing policy and practice impacts from social science research*, Submitted to Research Evaluation 21 January 2008

97. Merkx F, van der Weijden I, Oostveen A M, van den Besselaar P and Spaapen J (2007) *Evaluation of research in context: quick scan of an emerging field*. The Hague: Rathenau Instituut and EriC (Evaluating Research in Context). http://www.rathenau.nl/showpageBreed.asp?steID=1&ID=2963.

98. Michelson E S (2006) *Approaches to research and development performance assessment in the United States: an analysis of recent evaluation trends*. *Science and Public Policy* 33(8): 546-560.

99. Milner S J, Bailey C, Deans J and Pettigrew D (2005) Integrated impact assessment in the UK: use, efficacy and future development. *Environmental Impact Assessment Review* 25(1): 47-61.

100. Molas-Gallart J. and Tang P (2007) Policy and practice impacts of ESRC funded research: case study of the ESRC Centre for Business Research. Swindon: Economic and Social Research Council.

101. Molas-Gallart J., Tang P. and Morrow S (2000) Assessing the non-academic impact of grant-funded socio-economic research: results from a pilot study. *Research Evaluation* 9(3): 171-182.

102. National Audit Office (2003) *Getting the evidence: using research in policy making*, HC 586-I, session 2002-2003. London: Stationery Office. http://www.nao.org.uk/publications/nao_reports/02-03/0203586-i.pdf.

103. National Audit Office (2003) *An international review on governments' research procurement strategies. A paper in support of Getting the evidence: Using research in policy making*, HC 586-II, session 2002-2003. London: Stationery Office. http://www.nao.org.uk/publications/nao_reports/02-03/0203586-ii.pdf.

104. National Health Service SDO (2007) The impact of the NHS Service Delivery and Organisation Research and Development Programme 2001-2006. London: National Co-ordinating Centre for the NIHR Service Delivery and Organisation Programme (NCCSDO). http://www.sdo.lshtm.ac.uk/files/adhoc/impact-study-full.pdf.

105. National Research Council (2007) *Prospective evaluation of applied energy research and development at DOE (Phase Two)*. Washington, DC: National Academies Press. http://books.nap.edu/catalog.php?record_id=11806.

106. Nederhof A (2006) Bibliometric monitoring of research performance in the social sciences and the humanities: a review. *Scientometrics* 66(1): 81-100.

107. Norton G W and Alwang J (1998) *Policy for plenty: measuring the benefits of policy-oriented social science research*, Impact Assessment Discussion Paper No. 6. Washington, DC: International Food Policy Research Institute. http://www.ifpri.org/impact/iadp06.pdf.

108. Nutley S, Percy-Smith J and Solesbury W (2003) *Models of research impact: a cross sector review of literature and practice*, Building Effective Research 4. London: Learning and Skills Research Centre. http://www.lsrc.ac.uk/publications/index.asp.

109. Office of Science and Innovation (2006) *Science review of the Department for Environment, Food and Rural Affairs*. London: Office of Science and Innovation. http://www.berr.gov.uk/files/file35751.pdf.

110. Orosz E (1994) The impact of social science research on health policy. *Social Science & Medicine* 39(9): 1287-1293.

111. Paarlberg R L (1999) External impact assessment of IFPRI's 2020 Vision for Food, Agriculture, and the Environment Initiative, Impact Assessment Discussion Paper No. 10. Washington, DC: International Food Policy Research Institute. http://www.ifpri.org/impact/iadp10.pdf.

112. Pang T, Sadana R, Hanney S, Bhutta Z A, Hyder A A and Simon J (2003) Knowledge for better health: a conceptual framework and foundation for research systems. *Bulletin of the World Health Organization* 81(11): 815-820.

113. Park D (1998) *A proposal for measuring the benefits of policy-oriented social science research*, Impact Assessment Discussion Paper No. 3. Washington, DC: International Food Policy Research Institute. http://www.ifpri.org/impact/iadp03.pdf.

114. Peckham S, Willmott M, Allen P, Anderson S and Goodwin N (forthcoming 2008) Assessing the impact of the NHS Service Delivery and Organisation Research Programme. Submitted to *Evidence & Policy*.

115. Percy-Smith J, Speller V and Nutley S (2006) *Evidence Informed Policy and Practice: a review of approaches used in health improvement in Scotland*. Edinburgh: NHS Health Scotland. http://www.healthscotland.com/documents/1318.aspx.

116. Pestieau C (2003) *Evaluating policy research*, Research Paper W22. Ottawa: Canadian Policy Research Networks. http://www.cprn.org/documents/24336_en.pdf.

117. Pielke R A (1995) Usable information for policy: an appraisal of the US Global Change Research Program. *Policy Sciences* 28(1): 39-77.
118. PREST (Policy Research in Engineering, Science and Technology) (2006) *Smart innovation: a practical guide to evaluating innovation programmes*. Manchester: University of Manchester, PREST. http://www.mbs.ac.uk/Research/engineeringpolicy/publications/documents/SMARTInnovation.pdf.

119. Research Council Economic Impact Group (2006) *Increasing the economic impact of research councils: advice to the Director General of Science and Innovation, DTI from the Research Council Economic Impact Group*, 06/1678. London: Department of Trade and Industry. http://www.berr.gov.uk/files/file32802.pdf.

120. Research Councils UK (2007) Increasing the economic impact of the research councils. Swindon: Research Councils UK. http://www.rcuk.ac.uk/cmsweb/downloads/rcuk/publications/ktactionplan.pdf.

121. Rider Smith D with Sutherland A (2002) *Institutionalizing IMPACT Orientation: building a performance approach that enhances the impact orientation of research organizations*. London: University of Greenwich, Natural Resources Institute. http://www.parcinfo.org/documents/PARC%20Library/Institutionalizing%20Impact%20Orientation.pdf.

122. Rigby E (2002) *Assessing impacts of social science research*. Paper presented at Social Science Research Impact: Mapping and Measurement, International Social Science Comparisons Workshop, Berlin-Brandenburgische Akademie der Wissenschaften, Berlin, 14-15 March 2002. Berlin. http://www.stage-research.net/SPSG/paris-issc/BerlinRigby.html.

123. Riggirozzi M P and Tuplin T (2004) *The influence of research on policy: MIMAP Philippines*. Ottawa: International Development Research Centre, Evaluation Unit. http://www.idrc.ca/en/ev-57585-201-1-DO_TOPIC.html.

124. Ryan J G (1999) *Assessing the impact of rice policy changes in Viet Nam and the contribution of policy research*, Impact Assessment Discussion Paper No. 8. Washington, DC: International Food Policy Research Institute. http://www.ifpri.org/impact/iadp08.pdf.

125. Ryan J G (2002) *Synthesis report of workshop on assessing the impact of policy-oriented social science research in Scheveningen, the Netherlands, November 12-13, 2001*, Impact Assessment Discussion Paper No. 15. Washington, DC: International Food Policy Research Institute. http://www.ifpri.org/impact/iadp15.pdf.

126. Ryan J G and Garrett J L (2003) *The impact of economic policy research: lessons on attribution and evaluation from IFPRI*, Impact Assessment Discussion Paper No. 20. Washington, DC: International Food Policy Research Institute. http://www.ifpri.org/impact/iadp20.pdf.
127. Sayce K and Norrish P (2006) *Perceptions and practice: an anthology of impact assessment experiences*. Wageningen: Technical Centre for Agricultural and Rural Cooperation. http://www.anancy.net/uploads/file_en/impact%20assessment.pdf.

128. Science Advisory Council (2007) *CSA response to SAC recommendations on the End to End Review of Science into Policy in Defra*, SAC (07) 02. London, Department for Environment, Food and Rural Affairs. http://www.defra.gov.uk/science/documents/papers/2007/SAC(07)02.pdf.

129. Science and Technology Committee (2006) *Research council support for knowledge transfer: government response to the Committee's third report of session 2005-06: fifth special report of session 2005-06*, HC 1653. London, Stationery Office. http://www.publications.parliament.uk/pa/cm200506/cmselect/cmsctech/1653/1653.pdf.

130. Scott A (2000) *The dissemination of the results of environmental research*, Environmental Issues Series No. 15, Office for Official Publications of the European Communities: Luxembourg. http://reports.eea.europa.eu/Environmental_issues_No_15/en.

131. Scott A, Steyn G, Geuna A, Brusoni S and Steinmueller E (2001) *The economic returns to basic research and the benefits of university-industry relationships: a literature review and update of findings*. Brighton: University of Sussex, SPRU-Science and Technology Policy Research. http://www.sussex.ac.uk/spru/documents/econreturnsost.pdf.

132. Smith V H (1998) *Measuring the benefits of social science research*, Impact Assessment Discussion Paper No. 2. Washington, DC: International Food Policy Research Institute. http://www.ifpri.org/impact/iadp02.pdf.

133. Soper B and Hanney S R (2007) Lessons from the evaluation of the UK's NHS R&D Implementation Methods Programme. *Implementation Science* 2(7). http://www.implementationscience.com/content/pdf/1748-5908-2-7.pdf.

134. Spilsbury M J and Kaimowitz D (2000) The influence of research and publications on conventional wisdom and policies affecting forests. *Unasylva* 51(203): 3-10. http://www.fao.org/docrep/x8080e/x8080e02.htm#P0_0.

135. Thiele G, Devaux A, Velasco C and Manrique K (2006) *Horizontal evaluation: stimulating social learning among peers*. ILAC Brief 13, Rome: Institutional Learning and Change.
http://www.cgiar-ilac.org/downloads/Briefs/ILAC_Brief13.pdf.

136. Tress B, Tress G and Fry G (2005) Integrative studies on rural landscapes: policy expectations and research practice. *Landscape and Urban Planning* 70(1-2): 177-191.

136a. Trostle J, Bronfman M and Langer A (1999) How do researchers influence decision-makers? Case studies of Mexican policies, *Health Policy and Planning*; 14 (2): 103-114

137. UK Evaluation Forum (2006) *Medical research: assessing the benefits to society*. London: The Academy of Medical Sciences, Medical Research Council and Wellcome Trust.
http://www.wellcome.ac.uk/assets/wtx031817.pdf.

138. Walter I, Nutley S and Davies H (2003) *Research impact : a cross sector review – literature review*. St Andrews: St Andrews University, Research Unit for Research Utilisation.
http://www.ruru.ac.uk/discussion_papers.html.

139. Walter I. Nutley S and Davies H (2004) *Assessing research impact: report of seminar, 15-16 January 2004*. St Andrews: University of St Andrews, Research Unit for Research Utilisation.
http://www.ruru.ac.uk/discussion_papers.html.

140. Warburton D (2006) *Evaluation of Your Health, Your Care, Your Say: an independent report commissioned by the Department of Health*. London: Department of Health. Available via
http://www.dh.gov.uk/en/Publicationsandstatistics/index.htm.

141. Weiss C H (2003) *Studying research utilization*, Paper prepared for IDRC meeting, Ottawa, March 24-25, 2003.
http://www.idrc.ca/en/ev-66465-201-1-DO_TOPIC.html.

142. Williams D and Rank A D (1998) Measuring the economic benefits of research and development: the current state of the art. *Research Evaluation* 7(1): 17-30.

143. Wooding S, Hanney S, Buxton M and Grant J (2004) *The returns from arthritis research: a report prepared by RAND Europe for the Arthritis Research Campaign. Volume 1: approach, analysis and recommendations*. Cambridge: RAND Europe.
http://www.rand.org/pubs/monographs/MG251/.

144. Wooding S, Nason E, Klautzer L, Rubin J, Hanney S and Grant J (2007) *Policy and*

practice impacts of research funded by the Economic and Social Research Council. a case study of the Future of Work programme, approach and analysis. Cambridge: RAND Europe. http://www.rand.org/pubs/technical_reports/2007/RAND_TR435.pdf.

145. World Bank (2004) *Monitoring and evaluation: some tools, methods and approaches*. Washington, DC, The World Bank. Available via http://www.worldbank.org/ieg/.

146. Zillberman D and Heiman A (1999) *The value of economic research*, Impact Assessment Discussion Paper No. 7. Washington, DC: International Food Policy Research Institute. http://www.ifpri.org/impact/iadp07.pdf.

147S Aeserud, M., Lewin, S., Innvaer, S., Paulsen, EJ., Dahlgren, AT., Tommald, M., Duley, L., Zwarenstein, M. and Oxman, AD. (2005) 'Translating research into policy and practice in developing countries: a case study of magnesium sulphate for pre-eclampsia' *BMC Health Services Research* 5: 68-81

148S Armstrong, R., Waters, W., Crockett, B. and Keleher, H. (2007) 'The nature of evidence resources and knowledge translation for health promotion practitioners' *Health Promotion International* 22(3): 254-260

149S Australian Government (2005) *Research Quality Framework: Assessing the quality and impact of research in Australia.* Issues Paper [online]. Available from: http://www.dest.gov.au/NR/rdonlyres/E32ECC65-05C0-4041-A2B8-75ADEC69E159/4467/rqf_issuespaper.pdf [Accessed 15/12/2008]

150S Belkhodja, O., Amara, N., Landry, R., Ouimet, M. (2007) 'The extent and organisational determinants of research utilisation in Canadian Health Services Organisations' *Science Communication* 28: 377-417

151S Brutscher, PB., Wooding, S., Grant, J. (2008) *Health Research Evaluation Frameworks.* Technical Report [online]. Available from: http://www.rand.org/pubs/technical_reports/2008/RAND_TR629.pdf [Accessed 17/12/2008]

152S CAHS (2009) *Making an Impact: A preferred framework and indicators to measure returns on investment in health research.* Assessment Report [online]. Available from: http://www.cahs-acss.ca/e/pdfs/ROI_FullReport.pdf [Accessed 19/01/2009]

153S CGIAR (2008) *Impact Assessment of Policy-Oriented Research in the CGIAR: Evidence and Insights from Case Studies* [online]. Available from:

http://www.sciencecouncil.cgiar.org/fileadmin/user_upload/sciencecouncil/Impact_Assessment/PORIA2008_pr_4_F_l-r_3_.pdf [Accessed 12/12/2008]

154S Cook, J., Dickenson, S., Heuman, D. and white, G. (2008) *Making the most of evaluation.* SQW Consulting Viewpoints Issue 3 [online]. Available from: http://www.sqw.co.uk/file_download/147 [Accessed 06/01/2009]

155S COSEPUP (1999) *Evaluating Federal Research Programmes: Research and the Government Performance Results Act. The Summary and Statement of the Problem.* Washington, The National Academies Press.

156S Court, J. and Young, J. (2003) *Bridging Research and Policy: Insights from 50 Case studies.* ODI Working Paper 213. London, Overseas Development Institute.

157S CPHI (2002) *Tools and knowledge exchange: Scanning best practices in policy research*. Canadian Population Health Initiative Working Paper [online]. Available from: http://secure.cihi.ca/cihiweb/en/downloads/cphi_policy_practices_e.pdf [Accessed 16/12/09]

158S Crewe, E. and Young, J. (2002) *Bridging research and policy: Context, evidence and links* ODI Working Paper 173. London, Overseas Development Institute.

159S Donovan, C. (2008) 'The Australian Research Quality Framework: A live experiment in capturing the social, economic, environmental and cultural returns of publicly-funded research' *New Directions for Evaluation* 118: 47-60

160S Donovan, C. (2007) 'The qualitative future of research evaluation' *Science and public policy* 34(8): 585-597

161S Engel-Cox, JA., Van Houten, B., Phelps, J., Rose, SW. (2008) 'Conceptual model of comprehensive research metrics for improved human health and environment' *Environmental Health Perspectives* 5: 583-592

162S English, J. and Renkow, M. (2007) *The impacts of IFPRI's Global Research Program on the Sustainable Development of Less Favoured Areas.* Discussion Paper 26. Washington, IFPRI.

163S FRST (2003) *Natural Ecosystem Research Evaluation Report* [online]. Available from: http://www.frst.govt.nz/files/Natural_Ecosystem_Research_Evaluation.pdf [Accessed 11/12/2008]

164S Garrett, JL. and Islam, Y. (1998) *Policy research and the policy process: Do the twain ever meet?* IIED Gatekeeper Series 74 [online]. Available from: http://www.iied.org/pubs/pdfs/6138IIED.pdf [Accessed 11/12/2008]

165S Gilson, L. and McIntyre, D. (2008) 'The interface between research and policy: Experience from South Africa' *Social science and medicine* 67: 748-759

166S Gunderson, M. (2007) 'How academic research shapes labor and social policy' *J. Labor Res.* 28: 573-590

167S Hayward, SF. (2006) 'Environmental Science and Public Policy' *Social Research* 73(3): 891-914

168S ISSUES-EPSRC Knowledge Exchange (2008) *Engaging stakeholders and disseminating results*. ISSUES Project Workshop Report [online]. Available from: http://www.urbansustainabilityexchange.org.uk/media/KE%20Questions%20for%20Researchers%20from%20ISSUES%20Event%2010Nov08.pdf [Accessed 13/01/2009].

169S Kalucy, L., Bowers, EJ., McIntyre, E. and Reed, R. (2007) *Primary Health Care Research Impact Project: Final Report Stage I* [online]. Available from: http://www.phcris.org.au/phplib/filedownload.php?file=/elib/lib/downloaded_files/publications/pdfs/phcris_pub_3338.pdf [Accessed 11/01/2009]

170S Kanninen, S. and Lemola, T. (2006) *Methods for evaluating the impact of basic research funding: An analysis of recent international activity.* Helsinki, Academy of Finland.

171S Number not used as found to be duplicate of reference 79a.

172S Kothari, A., MacLean, L. and Edwards, N. (2009) 'Increasing capacity for knowledge translation: understanding how some researchers engage policy-makers' *Evidence and Policy* 5(1): 33-51

173S Kwan, P., Johnston, J., Fung, AYK., Chong, DSY., Collins, RA. and Lo, SV. (2007) 'A systematic evaluation of payback of publicly-funded health and health services research in Hong Kong' *BMC Health Services Research* 7: 121-131

174S Landry, R., Lamari, M., Amara, N. (2003) 'The extent and determinants of the utilisation of university research in Government agencies' *Public Administration Review* 63(2): 192-205

175S Lavis, JN. (2006) 'Research, public policy-making and knowledge-translation processes: Canadian efforts to build bridges' *The Journal of Continuing Education in the Health Professions* 26: 37-45

176S Lavis, JN., Oxman, AD., Moynihan, R. and Paulsen, E. (2008) 'Evidence-informed health policy I – Synthesis of findings from a multi-method study of organisations that support the use of research evidence' *Implementation Science* 3: 53-60

177S Lawton, JH. (2007) 'Ecology, politics and policy' *Journal of Applied Ecology* 44: 465-474

178S O'Leary, G., Boyle, R., Donlon, B. and Sheils, L. (2008) 'Providing a framework for accountability and learning in environmental research' *Administration* 55(4): 159-170

179S Lewison, G. (2003) 'Beyond outputs: new measures of biomedical research impact' *Aslib Proceedings* 55(12): 32-42

180S Lewison, G. (2005) 'Beyond SCI citations – new ways to evaluate research' *Current science* 89(9): 1524-1530

181S Lillis, DA. (2000) 'Towards a new science evaluation system for New Zealand' *Research Evaluation* 8(2): 145-150

182S Martens, PJ. and Roos, NP. (2005) 'When health services researchers and policy-makers interact: tales from tectonic plates' *Healthcare Policy* 1(1): 72-84

183S Mcwilliam, CL., Stewart, M., Brown, JB., Feightner, J., Rosenberg, M., Gutman, G., Penning, M., Stewart, M., Tamblyn, R. and Morfitt, G. (2003) 'Promoting evidence-based health policy, programming and practice for seniors: Lessons from a National Knowledge Transfer Project' *Canadian Journal on Ageing* 22(4): 415-430

184S Mostert, E. and Raadgever, GT. (2008) 'Seven rules for researchers to increase their impact on the policy process' *Earth Syst. Sci.* 12: 1087-1096

185S Nason, E., Janta, B., Hastings, G., Hanney, S., O'Driscoll, M. and Wooding, S. (2008) *Health Research – Making an impact: The economic and social benefits of HRB-funded research* [online]. Available from: http://www.hrb.ie/uploads/tx_hrbpublications/Health_Reseach-Making_an_Impact.pdf [Accessed 15/12/2008]

186S NERC (2008) *Economic impact reporting framework* [online]. Available from: http://www.nerc.ac.uk/about/perform/documents/eirf0708.pdf [Accessed 22/11/2008].

187S MORST (2004) *Evaluation of the Environmental Output Class*. Wellington, New Zealand Ministry of Research, Science and Technology.

188S Norse, D. and Tschirley, JB. (2000) 'Links between science and policy-making' *Agriculture, Ecosystems and Environment* 82: 15-26

189S Nutley, S. (2003) *Increasing research impact: early reflections from the ESRC Evidence Network*. RURU Working Paper 16. London, ESRC UK Centre for Evidence Based Policy and Practice.

190S Omamo, SW. (2004) *Bridging research, policy and practice in African agriculture.* IFPRI Discussion Paper. Washington, IFPRI.

191S Ootwijn, W., Hanney, S., Ligtvoet, A., Hoorens, S., Wooding, S., Grant, J., Buxton, M. and Bouter, L. (2008) 'Assessing the impact of health technology assessment in the Netherlands' *International Journal of Technology Assessment in Health Care* 24: 259-269

192S Ormerod, SJ., Barlow, ND., Marshall, EJP. And Kerby, G. (2002) 'The uptake of applied ecology' *Journal of Applied Ecology* 39: 1-7

193S Otronen, M. (2004) ‘The evaluation of social relevance in an environmental research programme’ *Research Evaluation* 13(1): 43-50

194S Owens, S. (2005) ‘Making a difference? Some perspectives on environmental research and policy’ *Trans. Inst. Br. Geogr.* 30: 287-292

195S Petticrew, M., Whitehead, M., Macintyre, SJ., Graham, H. and Egan, M. (2004) ‘Evidence for public health policy on inequalities 1: The reality according to policy-makers’ *J. Epidemiol. Community Health* 58: 811-816

196S Powers, A., Rose, SW., Bernichon, T., Engel-Cox, J., Sill, M. and Versendaal, K. (2006) *Program Assessments and Evaluation for NIEHS: Work Assignment 1: Literature Reviews on Research Impact Assessment Methods and Concepts*. Arlington, Centers for Public Health Research and Evaluation.

197S Raitzer, D. and Ryan, JG. (2008) ‘State of the art in impact assessment of policy-oriented international agricultural research’ *Evidence and Policy* 4(1): 5-30

198S Riley, J. (1998) ‘Guidelines for an Assessment Method for the Optimum Uptake of Research’ *Journal of Sustainable Agriculture* 12(2/3): 99-117

199S Ryan, J. (2002) ‘Assessing the impact of food policy research: rice trade policies in Vietnam’ *Food Policy* 27: 1-29

200S Spangaro, J. (2007) ‘Eleven obstacles to translating research into policy on gender-based violence’ *Evidence and Policy* 3(4): 553-566

201S Spilsbury, MJ. and Nasi, R. (2006) ‘The interface of policy research and the policy development process: challenges posed to the forestry community’ *Forest policy and economics* 8: 193-205

202S Trochim, WM., Marcus, SE., Masse, LC., Moser, RP. and Weld, PC. (2008) ‘The evaluation of large research initiatives: a participatory integrative mixed methods approach’ *American Journal of Evaluation* 29: 8-28

203S Van der Meulen, B. and Rip, A. (2000) ‘Evaluation of societal quality of public sector research in the Netherlands’ *Research Evaluation* 8(1): 11-25

204S Vogel, I. (2001) ‘An online assessment tool for research information: some preliminary concepts’ *Information Development* 17: 111-114

205S Willison, DJ. and MacLead, SM. (1999) ‘The role of research evidence in pharmaceutical policy-making: evidence when necessary but not necessarily evidence’ *Journal of evaluation in clinical practice* 5(2): 243-249

206S Wimpenny, P., Johnson, N. Walter, I. and Wilkinson, J. (2008) 'Tracing and identifying the impact of evidence – use of a Modified Pipeline Model' *Worldviews on Evidence-Based Nursing* 5(1): 3-12

207S Wooding, S., Hanney, S. and Grant, J. (2005) 'Payback arising from research funding: Evaluation of the Arthritis Research Campaign' *Rheumatology* 44: 1145-1156

208S Mollas-Gallart, J. and Tang, P. (2007) *Report of the ESRC Impact Evaluation Methods Workshop* [online]. Available from: http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/Images/ESRC%20Impact%20Evaluation%20Methods%20Workshop_tcm6-24818.pdf [Accessed 15/12/2008]

209S Morton, S., Nutley, S. and Jung, T. (2008) *The impact of the social sciences on public policy and the impact of evidence-based policy on the social sciences* [online]. Available from: http://www.ruru.ac.uk/events.html [Accessed 15/12/2008]

12 Appendices

12.1 Literature Review methods

The literature review's searches and analysis were conducted drawing on principles from systematic reviewing. The search process covered literature databases, internet searching, hand searching important journals, citation follow-up and personal recommendations. The material was read and reviewed for information relevant to the project's research questions.

Whilst formal quality appraisal methods were not employed on the sources identified, the majority of sources had either been peer reviewed as part of formal journal peer-review processes, or represented the outcomes of large well-funded investigations (e.g. by CAHS, CGIAR). Conclusions drawn from sources that did not fall into one of these categories were included only after careful consideration of the methods employed in the research studies in question and with an awareness of potential limitations.

1. Database searches

A number of databases were searched in order to identify wide-ranging literature, covering not just journals but also other published literature (monographs, conference proceedings and papers), and grey literature. Eleven key databases were searched by PSI's information scientist: COPAC, Enviroline (LexisNexis), Environmental Sciences & Pollution Management (CSA), Geobase, Institutional Repository search (Intute), Pollution Abstracts, SCOPUS, Social Policy & Practice, Web of Knowledge (Science Citation Index; Social Sciences Citation Index), Zetoc, and FRANCIS.

The search was limited to references published between 1997 and 2009, and initially limited to English language references.

Before undertaking the database searches, potential search terms were identified under each of four themes; research impact, evaluation methods, policy-making and environment.

- *Research impact*: 'research impact', 'research', 'research utili?ation', 'impact', 'influence', 'benefit*', 'outcome', 'payback'
- *Evaluation methods*: 'Evaluat*', 'measure*', 'assess*', 'method', 'quantif*', 'synthesis', 'review', 'model'
- *Policy making*: 'policy', 'decision?making', 'innovation', 'implement*', 'strategy', 'regulat*', 'payback', 'research uptake', 'uptake'
- *Environment:* 'environment*', 'pollut*'

N.B * = a truncated term, ? = a wildcard

Whilst the relevance of references from the health literature was apparent from the original Defra review, health-related terms were not included in the searches. Including health risked broadening up the searches too widely given available resources, and the overall review question was primarily interested in the impact of *environmental* research. Indirectly retrieved health-related references which otherwise met relevance criteria were not, however, excluded during initial screening.

The search terms above were combined within and across themes using Boolean operators i.e. OR within themes, AND or NOT across themes. Where available, and appropriate, proximity searching was used (e.g. in Web of Knowledge), and SAME to search for terms from different themes occurring in the same sentence, to retrieve any records with a relationship between those terms. Due to the general nature of the search terms, but the complexity of the relationships between the concepts being searched (and to avoid possible non-retrieval of potentially relevant references), searches sometimes retrieved high numbers of references. These were then reduced by further refining (e.g. by selecting groups from different subject areas if the functionality allowed this), and screening.

As database functionality differs, it is not possible to conduct identical searches in different databases. The capacity to search various fields varies; different combinations of controlled vocabulary terminology and free text terms need to be used.

As the title field is normally one of the keyword fields for database searches, title words in documents included in the Defra review bibliography were taken into account.

A further set of **cited reference searches** were also carried out in Web of Knowledge. From the bibliography of the Defra review, 29 particularly valuable references were selected for these cited reference searches.

Articles identified by these different searches were subsequently selected for inclusion based on a closer examination of the abstract (or full paper where an abstract was not available), and consideration of how each source met the core review questions.

2. Web searches

This involved a more creative search of 42 key organizations, including UK government departments, Select Committees, evaluation societies and research councils. The websites of the following organizations were searched: AHRC (Arts and Humanities Research Council), Australian Research Council, Biotechnology and Biological Sciences Research Council (BBSRC), Canadian Health Services Research Foundation, Canadian Environmental Assessment Research Council, CGIAR (Consultative Group on International Agriculture Research), UK Department of Health, Defra, Engineering and Physical Sciences Research Council (EPSRC), ERFF (Environmental Research Funders' Forum), European Environment Agency, Economic and Social Research Council (ESRC), European Evaluation Society, Ecologic, Eldis, The European Policy Evaluation Consortium, Finnish Environment Institute (SYKE), Government Accountability Office, Health Economics Research Group (HERG),

International Development Research Centre (IDRC), Institute for Development Studies (IDS), Institute for International Environment and Development (IIED), Institute for European Environmental Policy (IEEP), International Food Policy Research Institute (IFPRI), National Academies Press, Natural Environment Research Council (NERC), Natural Sciences and Engineering Research Council of Canada (NSERC), National Sciences Foundation (NSF), NHS Service Delivery Organisation (NHS SDO), Overseas Development Institute (ODI), PREST (Centre for Science and Technology Policy and Management Research), The Peer Initiative, REPP (Research Evaluation and Policy Project), Research Unit for Research Utilisation (RURU), Social Care Institute for Excellence (SCIE), Science and Technology Policy Research (SPRU), Social Policy Evaluation and Research Committee, RAND Europe, Stockholm Environment Institute (SEI), Technopolis, UK Evaluation Society (UKES), World Bank.

The database and web search were conducted in December 2008.

3. Citation tracking and expert contacts

This involved following up references and contacts emerging from documents identified through the initial stages of the search. Hand-searching key journals in the field (Evidence and Policy, Evaluation Research, Research Evaluation, and Social and Public Policy) also identified some additional papers. The project team were also contacted to see if they were aware of any additional sources.

4. Searching of non-English language literature

Efforts were also undertaken to access and review non-English language publications through drawing on the language skills possessed by the research consortium, including French and Dutch. Given the limited resources available for this, four different approaches were taken in order to gain an idea of how much non-English language literature may be available that had not already been translated into English:

- The initial database search included searches of databases such as FRANCIS which were more likely to access European literature in this area.

- The initial database searches were repeated in Web of Knowledge, this time selecting for French language publications. This retrieved ten references, two of which were of relevance. These had, however, already been translated into English and therefore already detected and reviewed in the English language searches.

- The search terms were translated into French and Dutch and appropriate databases searched to try to identify relevant articles. This detected a limited number of articles in Dutch which, again had already been translated into English.

- The survey of SKEP network members asked about any relevant articles they were aware of in their own countries. If written in languages other than those possessed by

the project team, SKEP members were asked to send any available English summaries of these documents.

12.2 Data extraction sheet template

This is the data extraction sheet template that was used to collate material of relevance to this project's objectives and research questions.

SKEP review data extraction sheet	
ID reference number (from PSI)	
Method of identification (e.g. database)	
Details of publication Author Title Source (journal/conference etc) Year/volume/pages Country of origin Language Institutional affiliation	
Field (e.g. environmental research)	
Type of paper (e.g. conceptual, empirical, descriptive)	
Research question/aim	
Study design	
Theory	
Analysis	
Findings/key points (relevant to this project's research question)	
Reason for inclusion: Updating - new material Expanding - non-English language - previously unknown material **Paper includes** (tick as appropriate) • An evaluation of the impact of research on policy • A reflective piece about evaluation of the impact of research on policy • A description of method(s) for evaluating impact of research on policy • Evaluation of effectiveness/value for money of methods • A review of methods for evaluating impact more generally • Others of relevance (give reason)	