STATUS: PENDING 20120531    OCLC #: 58832548
REQUEST DATE: 20120531 NEED BEFORE: 20120628    SOURCE: FSILLSTF
BORROWER: DOS    RECEIVE DATE:    DUE DATE:
RENEWAL REQ:    NEW DUE DATE:    SPCL MES:
LENDERS: UAB, *UW1, YQU, SFB, UBC

TITLE: Evidence & policy : a journal of research, debate and practice.
ISSN: 1744-2648
IMPRINT: Bristol, UK : Policy Press, 2005 9999
ARTICLE TITLE: is your evidence robust enough?
VOLUME: 1
ISSUE NUMBER: 1
ISSUE DATE: 2005
PAGES: 101-112
VERIFIED: WorldCat Desc: v. ;Type: Serial
SHIP TO: ATTN: INTERLIBRARY LOAN /Room 3239/U.S. Dept. of State /Ralph Bunche Library/2201 C Street, NW/Washington DC US 20520-2442
BILL TO: IFM ONLY
SHIP VIA: LIBRARY RATE
MAXCOST: IFM - 15.00
COPYRIGHT COMPLIANCE: CCL
FAX: 202-647-2971
EMAIL: library@state.gov
AFFILIATION: LVIS; PEDLINK
LOCATIONS: KIJ, QCL, ERL, UV0, LGG, WEA
BORROWING NOTES: *****Please use CONDITIONAL if your fee is higher than our maximum allowed.

PATRON: Manamperi, P
Is your evidence robust enough? Questions for policy makers and practitioners

Louise Shaxson

English This article examines the reasons we need evidence for policy, discusses where evidence is needed in the policy-making process, and the nature of the evidence base for strategy and policy. Working relationships between policy makers and their advisers are key: as policy makers come from a variety of backgrounds, developing a common language helps set discussions about the robustness of the evidence base on a sound footing. The article identifies five components of robustness, proposes a series of questions that could be used to address them and discusses the implications for the processes of policy making.

Français Cet article examine les raisons pour lesquelles nous avons besoin de preuves pour établir les politiques générales, où les preuves sont nécessaires dans le processus de prise de décision et la nature du principe de base sur preuves, pour les stratégies et politiques générales. Les relations de travail entre les dirigeants et leurs conseillers sont clé : étant donné que les dirigeants viennent d'une diversité de milieux socio-culturels, développer un langage commun permet de mettre en place des discussions, relatives à la solidité de la base de preuves. L'article identifie cinq composants de solidité, propose une série de questions, qui pourraient être utilisées pour aborder ces composants et examine les implications des processus de prise de décision.

Español Este artículo examina las razones por las que necesitamos evidencia para la política, examina dónde se necesita la evidencia en el proceso de la elaboración de la política a seguir; y la naturaleza de la base de la evidencia para la estrategia y la política. Las relaciones de trabajo entre los diseñadores de política y sus consejeros son claves: ya que los diseñadores de política proceden de diferentes formaciones, desarrollando una lengua común ayuda a establecer discusiones acerca de la solidez de la evidencia basada en una buena base. El artículo identifica cinco componentes de solidez, propone una serie de preguntas que se podrían usar para discutirlas y también discutir las implicaciones en los procesos de la elaboración de la política a seguir.

Key words evidence-based • policy • processes • robust

Key dates final submission 12 April 2004 • acceptance 12 May 2004
This discussion article began as a tool for government policy makers to help them communicate better with their advisers. I sat in meetings wondering whether or not they were all speaking the same language. If not, what were the implications for the policy-making process? What follows are my reflections on implementing an evidence-based approach to policy making, most of it gleaned from discussions with people working within, and as advisers to, the Department for Environment, Food and Rural Affairs (Defra). My conclusions reflect helpful comments from people with both policy and advisory backgrounds.

Why do we need evidence for policy?

In 1999, the White Paper, Modernising government (Cabinet Office, 1999), noted that government must “produce policies that really deal with problems; that are forward-looking and shaped by the evidence rather than a response to short-term pressures; that tackle causes not symptoms”. The accent on what works led to an emphasis on data rather than dogma, and on the development of various tools, such as the Regulatory Impact Assessment (www.cabinetoffice.gov.uk/regulation/index.asp), to improve the chain of causality between evidence and advice. The phrase “evidence-based rather than evidence-backed” was born.

However, is evidence-based policy making a distinct approach? Davies, referenced in Cabinet Office (nd) calls it “the integration of experience, judgement and expertise with the best available external evidence from systematic research”, and notes that it involves a shift away from opinion-based decision making towards decisions based on “the opinions and judgements of experts (that) constitute high quality valid and reliable evidence”. There are several unanswered questions in this statement, however. How do facts sit with opinions and judgements? What makes evidence high quality, valid and reliable? Some of these are addressed in the Office of Science and Technology’s (2000) Guidelines 2000; the processes of science advice, peer review and other high-level challenge functions have been greatly strengthened by the application of their principles. On the other hand, there is no firm guidance for policy makers as they discuss with their advisers which pieces of all the potentially relevant information they should use in preparing policy submissions. This discussion article was prepared to assist that process.

Evidence is a necessary, but not a sufficient, condition for any decision-making process; the ways that policies are developed, implemented, monitored and revised are always shaped by the wider social and political contexts. More specifically, we need evidence to:

- understand the policy environment and how it’s changing;
- appraise the likely effects of policy changes so we can choose between different policy options and subsequently assess their impacts;
- demonstrate the links between strategic direction, intended outcomes and policy objectives to show that there are clear lines of argument and evidence between what we are aiming for and what we are doing now;
- determine what we need to do to meet our strategic goals or intermediate objectives;
influence others so that they help us achieve our policy goals and take them through to delivery;
communicate the quality (breadth and depth) of our evidence base to meet the open government agenda.

What is evidence?
The Chambers English dictionary (CED, 1990) defines ‘evidence’ as “that which makes evident: means of proving an unknown or disputed fact: support for a belief: indication: information in a law case: testimony: a witness or witnesses collectively”. This is fine, but it tells us little about how to value the different types of evidence; and for policy purposes, I think Michael Harrison has come up with a better definition. In a draft of this article, he suggested:

Evidence for policy making is any information that helps to turn a department’s strategic priorities and other objectives into something concrete, manageable and achievable.

For Defra, and for government more widely, these strategic priorities are very broad indeed (see, for example, Defra, nd), and most of them overlap. Defining evidence in relation to how we make strategic priorities operational, emphasises the importance of internal processes that turn the soup of information into an evidence base upon which decision takers can make reasonable and defendable judgements.

However, the evidence base is dynamic. People’s understanding and interpretation change, new research results come in, we deepen our awareness of issues we previously thought had little connection, and we develop new ways of using and interpreting information that we already have. If the evidence base is changing, then so must the ways in which we manage, filter and use it for policy. Good evidence-based policy making is not simply about creating a vast database of everything and then cherry-picking the best, or most accessible or most immediately relevant information.

The remainder of the article discusses what constitutes good evidence-based policy making. However, it is important to note up front that the term ‘evidence base’ relates just as much to the processes of using information as it does to the quality of the information itself.

Figure 1 shows a very simple schema: if evidence-based policy making relates as much to processes of using information as it does to the quality of the information itself, then we can represent it as follows:

**Figure 1:** A simple schema of evidence-based policy making

- Policies that take good information ...
- Policies that take poor information ...
- ... and use it well
- ... and use it poorly
Obviously there is only one arrow that results in good evidence-based policies. The other policies may be evidence based, but not in the ways we would wish.

**Evidence and the policy-making process**

Evidence for policy making comes from a variety of sources at different times in the policy-making process, and departments use many approaches to filtering evidence and synthesising it into a coherent basis for policy decisions. Some of the literature on evidence-based policy making (see Garrett and Islam, 1998; Omamo, 2004) sets out different models and it appears that researchers are almost universally frustrated that every time they come across a new policy environment or policy question, none of the models appears to operate. Levitt (2003, p 14) has probably the best description of the process as a constantly shifting jigsaw, and her analysis demonstrates why trying to quantify the impact of research on policy is almost impossible (viz, Ryan and Garrett, 2003). However, discussions with Tony Taig have produced a far simpler model (Figure 2) which encompasses all those referenced earlier.

Evidence is needed both in response to time-sensitive requests for information, and to support longer-term strategy and policy development. Steven Were Omamo (2004, p 30) notes that “the issue is how to promote ‘evidence-readiness’ among inherently conservative and pragmatic policy makers and practitioners and ‘user-readiness’ among inherently abstraction-oriented researchers”, and suggests that neither individual policy makers nor individual researchers will ever be ready to “send and receive signals from one another”. While the former statement is undoubtedly true, I wonder if the general air of gloom in his second statement comes from a rather one-sided view of policy making.

Research results are undoubtedly important to policy, but research is essentially a long-term process. Tony Taig, in a personal communication, points out that much

---

**Figure 2: The flow of evidence in the policy process**

![Diagram showing the flow of evidence in the policy process](https://example.com/diagram.png)

Source: Tony Taig, personal communication
policy making happens in response to very short-term pressures, where the need for evidence can be nicely summarised as “please synthesise current knowledge to answer my pressing policy question”. For many policy makers, being able to produce options to place before a decision taker needs to happen in a few days or, in some cases, a few hours. Big, slowly evolving decisions that involve a good deal of strategic thinking are relatively rare, compared with the smaller (but no less important) issues that are constantly nipping at policy makers’ heels.

This creates an obvious tension. Buried on a page of Nutley et al’s paper (2002, p 4) is the question, “How can the need for rigour be balanced with the need for timely findings of practical relevance?”. Increasingly, I believe this is the key to improving evidence-based policy making. How do we correlate policy makers’ need for a rapid turnaround with the no less important need to provide evidence that is robust in the long term and that can be used to rigorously appraise policy options and for future monitoring and evaluation? Shortening the time frame does not mean we should cut corners: we still need to ensure that both policy makers and those who provide the evidence share a common understanding of the question, agree on the robustness of the evidence they need to answer it, and agree again how to interpret this to generate policy options.

This is becoming increasingly important for a science-heavy department such as Defra as it works through both the demands of government in relation to science, and the changing interface between science and society. Defra needs a very broad evidence base indeed, partly because the departmental goal is broad in scope, partly because of the wide range of policy objectives, and partly because of the diverse array of conflicting interests and perspectives on what constitutes ‘evidence’. The nature of the issues Defra is addressing inevitably requires questions to be framed in ways that bring together all the advisory disciplines – including the natural, engineering, social and economic sciences; statistics, legal advice and probably more – to develop practical management options. This means we need an approach that covers all possible evidence – opinions, judgements and analyses, as well as ‘hard’ facts (see Levitt, 2003, p 29) – and which is based on a common understanding of what we mean by ‘robust evidence’. But do policy makers and advisers have this common understanding? Are we talking the same language? What do we mean by ‘a robust evidence base for policy’? Figure 1 demonstrates that there is no absolute answer to this last question: in part, it depends on the way the question is asked (which shapes the information that emerges); but it also depends on the working relationships between policy teams and their advisers (which shapes how that information is used).

Building a consensus around ‘robust’ evidence

There is a well-established literature (see Marsland et al, 2001; Spencer et al, 2003) that analyses what we mean by ‘robustness’ and helps us think about how to improve the evidence base for policy making. While most of what follows is not new (see, for example, Crewe and Young, 2002; Court and Young, 2003; Spencer et al, 2003), I have extracted words from this literature that I believe make the most sense across the disciplines, resonate with people who are not completely familiar with the
concerns about quality of information, and have direct relevance to issues that are important to policy makers as they try to develop policy options under constant time pressure. It builds on work by Marsland et al (2001) to synthesise some of the more abstract academic analysis that has been done elsewhere and make it more accessible for policy makers and natural scientists, some of whom are unfamiliar with social science. The idea is to compose a language that can be shared by interdisciplinary teams of advisers and so that they can develop and present coherent messages and options to decision takers.

The framework is presented as a series of key questions around the five themes. There is no right or wrong answer for any of them, but as we develop a shared understanding of what we mean (‘we’ includes policy makers, advisers, disciplinary specialists, generalists, the public, other professional bodies), we begin to understand more about the policy goals and objectives, and about what others can bring to the policy-making process.

The four well-understood components of robustness are credibility, generalisability, reliability and objectivity (see Spencer et al, 2003). I suggest that there is a fifth – rootedness – that has direct and profound relevance for policy making. This is also defined as ‘authenticity’, and is examined to a degree in the qualitative literature, but has been ignored in the quantitative literature.

Credibility

This relates to the processes of analysing and synthesising information, often referred to as ‘internal validity’ in the quantitative literature. Credible evidence relies on a strong and clear line of argument, tried and tested analytical methods, analytical rigour throughout the processes of data collection and analysis, and on clear presentation of the conclusions. Key questions would include:

- How confident are we that these conclusions flow from this evidence? Would others come up with the same results?
- Does it matter whether a particular piece of evidence is sourced from expert or lay knowledge? How does it matter?
- Are our methods appropriate to the quality of the evidence? Does the evidence we are able to collect limit the question we are able to ask?
- Does the evidence make sense to the people we consulted, be they internal advisers, external experts or other stakeholders?
- Would involving our critics in gathering and analysing the evidence deepen its credibility?

Generalisability

Generalisability (also called ‘transferability’) refers to the way we make inferences. Can we take this evidence collected for a specific purpose and use it in a different context or to answer a different question? For some types of information, generalisability will refer primarily to sampling procedures; for others, it will be...
more about our understanding of context. It is particularly applicable when pilot studies precede a wider roll-out (see GCSRO 2003). Key questions could include:

- Are the findings widely applicable or context-specific? Is this to do with sampling techniques or with the broader framing of the issue and the policy question?
- Could it be that the arguments are generalisable even if the specific findings are not? Why?
- Which bits of the context matter or change the findings? Why?

**Reliability**

The standard literature on social research methods discusses reliability in a broad sense, treating it mainly in terms of the ability to replicate a study. For information from research, Departments formulate guidelines on procurement and review procedures to ensure that accepted best practice has been adhered to in sampling, analytical methods and other questions relating to the quality of research processes, for example. However, for policy making, the issues are different: we can’t replicate policies in a similar way. Extending the results from pilot areas raises considerations that are dealt with under generalisability, so I propose that for policy making, we should take a fairly narrow view of reliability, relating it to whether or not we can depend on the evidence for monitoring, evaluation or impact assessments. If this is overlooked at the beginning of policy work (that is, if baseline information is not collected before policies are implemented), then impact assessments will be meaningless. The government’s emphasis on ‘what works’ means planning for a lessons-learned approach by asking such questions as:

- Does the evidence form a sound basis for future monitoring, evaluation or impact assessment?
- Will the initial framing of the question hold up over time?
- What sort of contextual information will we need to monitor over time?
- For qualitative evidence in particular, is there a clear evidence trail that can be followed?

**Objectivity**

Again, there is an extensive literature on the various methods for reducing bias in the evidence base, but I think it useful to relate questions of bias specifically to the policy-making process. We need to explore bias in the evidence base for two reasons. First, because what we say is bounded by the assumptions we make or the restrictions we impose as we ask policy questions, and by the values we assign to different aspects of the evidence as we interpret it for policy (while social scientists are generally familiar with this argument, some policy makers and people with a natural science background may need help exploring it). The second reason for exploring bias (and I am grateful to Andy Stirling for this point) is that any weakness of a policy in relation to future events may lie in an implicit bias: one that we alone cannot currently see, but that conditions policy outcomes. We need to explore bias in the evidence
base, and deepen our understanding of how it conditions our interpretation of the evidence for policy, using questions such as:

- Have all possible techniques been used to remove bias from the evidence and from the analytical methods?
- Has any residual bias been acknowledged and accounted for in presenting the options to decision takers?
- Are we certain that the findings have been determined by the subjects and context of the inquiry rather than the biases, motivations and perspectives of the investigators, policy makers or decision takers?

**Rootedness**

The answers obtained from even the most well-documented body of knowledge will always reflect the particular questions that are asked, the way they have been posed and the use to which the answers are likely to be put. If questions are narrowly defined by those with a narrow, specialist knowledge of the subject, they will lead to narrow answers: the relationship between the evidence we have gathered and the conclusions we draw is inevitably sensitive to the processes we have used.

Rootedness is about more than context, process, bias and the quality of information. Rather, it is about understanding the nuance of the evidence, exploring assumptions with an open mind, encouraging others to question the status quo as we see it, and thinking about who uses what evidence for what purpose. We can address this by asking:

- Different evidence may have different meaning and value to different stakeholders. Have all the viewpoints been negotiated, both in framing the question and gathering the evidence?
- Does the question truly represent the fullness of the issue or are there other aspects that could and should be explored?
- What is the history of the evidence? How does this affect the question?
- Have the assumptions been examined by people with different specialist knowledge? Have the different implications been explored?
- Do the results stimulate action and empower people to act? Who is empowered? How are they empowered?

A summary question, adapted from Marsland et al (2001) is this: could the initial question, the evidence, the analytical methods, the conclusions and subsequent action be confirmed, refuted, explained or enriched by information from other sources?

**Taking a proportional approach**

Taking a proportional approach means screening the evidence (as well as the processes of gathering, analysing, interpreting, prioritising, and communicating it) against the five components of robustness; deciding which ones are important for each particular policy question, and working out how to achieve them given the inevitable time
and resource constraints. It does not mean waving this article at colleagues and point scoring for credibility or reliability in a particular part of the evidence base. It means sensitivity analysis, not only within the standard numerical framework to see what happens to costs when a parameter is changed, but also by relaxing some of the softer assumptions and then following the line of argument through to the end. It means deciding which aspects of robustness are important in which situation and why, and then working out how to achieve them given the inevitable time and resource constraints. It means cultivating an interdisciplinary approach to the evidence base, and working out how to foster these sorts of working relationships.

It is fairly clear that what I have separated as five components of robustness are all linked to each other to varying degrees. This is partly because of the complexity of the evidence base and partly because there is a circular relationship between question, information and analytical method. It is obviously impossible for all evidence in all policies to be completely robust all of the time. There will be times when different characteristics will be key: the importance of each characteristic may rise or fall during the policy-making process and in proportion to the perceived risk or the anticipated impact of a policy.

Next steps

This article is very much a work in progress. While the wording of the questions has been developed and tested with economists, social researchers and natural scientists, some policy makers have asked that the questions be made even simpler, more immediately relevant to the way they work and their short-term policy-making horizons. We will do this in the near future, by developing an easy reference system that helps policy teams think about who can best help them answer some of these complex questions about robustness.

Getting back to the first statement from the White Paper, Modernising government; we do need evidence that supports policy making in response to short-term pressures. We need to recognise that these pressures exist and greatly shape the way policy makers work. However, at the same time, we need to ensure that the questions are relevant to the entire evidence base, including the evidence for strategy and longer-term policy options. My hypothesis is that we can do this by working through the process of developing an operational strategy, establishing a clear line of argument and evidence between long-term strategy, medium-term intended outcomes and short-term policy outputs. At the same time, we need to agree what are the indicators of success. Given the overlapping nature of many of these strategic priorities, we need to also agree who ‘owns’ the indicators at each level; that is, who is responsible for delivering the outputs and who needs to ensure that aggregating the outputs leads you in the direction of the intended policy outcome. At all levels, we can test our statements for robustness, using the framework outlined in this article.

We can and should be refining the tools to help this process. My experience with logical frameworks shows that a group of people with very disparate backgrounds and experience can be drawn together around a mutual understanding of what the goals are and how each can contribute to their achievement. There may well be other tools, but they will need two defining characteristics. First, they need to offer
enough flexibility to encourage Omamo’s ‘user-readiness’ in researchers by encouraging them to focus on the real short-term pressures that policy makers face. At the same time, the tools must let researchers encourage policy makers to lift their eyes from their desks so that they are better able to understand how emerging evidence can be incorporated into their longer-term policy and strategy plans (Omamo’s ‘evidence-readiness’). Second, but perhaps more importantly, these tools must help us turn our strategic priorities into something concrete, manageable and achievable. They need to ensure that the evidence base for policy emerges from the evidence that underpins strategy.

Finally (and I owe this insight to John Holmes), developing this dialogue between policy maker and adviser in a spirit of open enquiry will result in a stimulating and innovative process. This will ensure that the issues are well explored, and that all the various perspectives enable questions to be well focused and the answers to be effectively interpreted and applied to policy.

**Note**

This is a discussion article – not a statement of Defra policy or processes. Various people in Defra and other organisations offered valuable comments and suggestions on previous drafts. I would particularly like to thank Michael Harrison, Christian Hudson, Miles Parker, Andy Stirling, Lindsey Poole, Andy Dobson, Faith Culshaw and John Holmes. Some of their suggestions are specifically acknowledged in the final text, but all errors and omissions are mine alone.

**References**


Defra (nd) *Defra’s aim and objectives* (www.defra.gov.uk/corporate/aims/aim.htm).


