

SEWPS

SPRU Electronic Working Paper Series

Paper No. 146

Improving the Evidence Base for Energy Policy: The Role of Systematic Reviews

Steve Sorrell
(SPRU)

February 2006

US

University of Sussex

SPRU – Science & Technology Policy Research

The Freeman Centre, University of Sussex,
Falmer, Brighton BN1 9QE, UK
Tel: +44 (0) 1273 877067
E-mail: s.r.sorrell@sussex.ac.uk
<http://www.sussex.ac.uk/spru/>

Improving the Evidence Base for Energy Policy: The Role of Systematic Reviews

February 2006

Steve Sorrell

Research Fellow

Sussex Energy Group

SPRU (Science & Technology Policy Research)

Freeman Centre

University of Sussex

Falmer,

Brighton,

BN1 9QE

Tel: +44 1273 877067

Fax: +44 1273 685865

Email: s.r.sorrell@sussex.ac.uk

Abstract

The concept of Evidence Based Policy and Practice (EBPP) has gained increasing prominence in the UK over the last ten years and now plays a dominant role in a number of policy areas, including healthcare, education, social work, criminal justice and urban regeneration. But despite this substantial, influential and growing activity, the concept remains largely unknown to policymakers and researchers within the energy field. This paper proposes a definition of EBPP, identifies its key features and examines the potential role of **systematic reviews** of evidence in a particular area of policy. It summarises the methods through which systematic reviews are achieved; discusses their advantages and limitations; identifies the particular challenges they face in the energy policy area; and assesses whether and to what extent they can usefully be applied to contemporary energy policy questions. The concept is illustrated with reference to a proposed review of evidence for a ‘rebound effect’ from improved energy efficiency. The paper concludes that systematic reviews may only be appropriate for a subset of energy policy questions and that research-funding priorities may need to change if they use is to become more widespread.

Keywords

Evidence based policy and practice; systematic reviews; research synthesis.

1. Introduction

From relatively small beginnings within the medical field, the concept of *Evidence Based Policy and Practice (EBPP)* has gained increasing prominence in the UK over the last fifteen years (Solesbury, 2001). An infrastructure now exists for developing and promoting evidence-based medicine and healthcare, both in the UK and abroad,ⁱ and the concept has been adopted by other professional and policy areas, including education, social work, criminal justice and urban regeneration (Davies, Nutley et al., 2000).ⁱⁱ The UK Economic and Social Research Council has established a network for conducting evidence-based reviews in the social sciences and is funding a Centre for EBPP at Queen Mary College, University of London. The concept informed the 1999 White Paper on Modernising Government, together with a number of subsequent UK government publications (Bullock, Mountford et al., 2001), and is central to the remit of the Cabinet Office. But despite this substantial, influential and growing activity, EBPP appears to be largely unknown to policymakers and researchers within the energy field (Gross, 2005).

This paper aims to introduce the concept of EBPP to energy policy researchers and evaluate its potential contribution to research and practice in this area.ⁱⁱⁱ Specifically, the paper seeks to:

- propose a definition of EBPP and identify its key features;
- introduce the concept of a **systematic review** of evidence in a particular area of policy or practice, and summarise the methods through which this is achieved;
- discuss the advantages and limitations of systematic reviews and the particular challenges they face in the energy policy area; and
- assess whether and to what extent the systematic review methodology can usefully be applied to contemporary energy policy questions.

The following section provides an introduction to EBPP, identifies its different elements and highlights the disputed nature of ‘evidence’ in different policy areas. Section 3 introduces a systematic review of evidence and summarises the stages and methods through which this is achieved. Section 4 highlights the limitations of systematic reviews, focusing upon a number of interrelated ‘biases’ to which they are prone. Section 5 discusses the challenges of applying systematic reviews to energy policy questions and illustrates this with reference to an example. Section 6 concludes.

ⁱ Including the Cochrane Collaboration, the National Institute for Clinical Excellence, the Health Development Agency and the NHS Centre for Reviews and Dissemination.

ⁱⁱ These have established parallel initiatives such as the Centre for Evidence Informed Education Policy and Practice (EPPI) at the University of London (Davies, Huw et al., 2000)

ⁱⁱⁱ The context for this paper is the [Technology and Policy Assessment](#) (TPA) function of the new [UK Energy Research Centre](#) (UKERC). The Imperial College Centre for Energy Policy and Technology (ICCEPT) leads the TPA in collaboration with the SPRU Energy Group at the University of Sussex. The TPA seeks to conduct two ‘evidence based’ assessments per year of important questions within UK energy policy. The first two assessments are on *The Costs and Impacts of Intermittency in Electricity Generation* (conducted by ICCEPT) and *The Evidence for a Rebound Effect from Improved Energy Efficiency* (conducted by SPRU). While the TPA seeks to draw upon EBPP ideas and techniques, it is not committed to the full application of the systematic review methodology. The present paper is based upon an earlier TPA report which examined the methodological issues in more detail (Sorrell, 2005).

2. What is Evidence Based Policy and Practice?

The rhetoric of EBPP pervades many areas of policy and practice within the UK. As the title suggests, its influence extends from low-level issues of professional practice, such as classroom techniques to encourage pupils with behavioural difficulties (Evans and Benefield, 2001), to high-level policy issues, such as the relationship between poverty, income inequality and violence (Hsieh and Pugh, 1993). But as with most such concepts, the precise meaning of ‘evidence-based’ is contested.

EBPP has been defined as “the integration of experience, judgement and expertise with the best available external evidence from systematic research” (Davies, 1999). Hence, EBPP implies striking a balance between formal research evidence and professional judgement. Davies and Nutley (2002) argue that for evidence to have a greater impact, there must be agreement as to what counts as evidence in what circumstances, a strategic approach to the creation of evidence in priority areas, systematic efforts to accumulate and synthesise evidence, effective dissemination of evidence to where it is most needed and initiatives to ensure better utilisation of evidence. But while each of these may be a necessary requirement for EBPP, the first presents a particular challenge. Most applications of EBPP have prioritised scientific research over other forms of evidence and given greater weight to some research methodologies than to others. Box 1 identifies four general types of research methodology, listing them in descending order of their weighting in EBPP.

Box 1 Different types of research evidence

- *Experimental and quasi-experimental evidence* is exemplified by randomised controlled trials in medicine, but also includes controlled before-and-after studies and various types of matched comparison that can be applied to policy evaluation and broader social scientific questions. This type of study potentially provides reliable evidence of the causal effect of different mechanisms by explicitly controlling for the effect of different variables.
- *Survey evidence* provides an alternative technique for exploring causal hypotheses, either through surveys conducted by the researcher or through the econometric analysis of existing data. In addition, survey data can provide valuable information about the nature, size, frequency and distribution of a particular variable or problem.
- *Modelling evidence* covers a variety of approaches to analysing the operation and consequences of different mechanisms using a simplified mathematical model. Like all theories, these models abstract from real-world complexities and focus on key mechanisms, either conceptually or by combining theoretical assumptions with empirical data.
- *Qualitative evidence* includes a variety of techniques for obtaining information regarding the opinions, attitudes and perceptions of individuals and groups in different contexts. Examples include case studies, participant observation and focus groups. Frequently, qualitative evidence will be combined with quantitative (e.g. survey) evidence to improve understanding about how and why a policy or mechanism works and under what conditions. Hence, rather than distinguishing between qualitative and quantitative approaches, it is perhaps more useful to compare those approaches that seek to establish *causal* inferences, whether through quantitative methods, qualitative methods or both, as compared to those which seek to *interpret* the experience of individuals and identify the meanings which those experiences hold.^{iv} The former tends to be associated with economics, political science and social psychology, while the latter tends to be associated with sociology and social anthropology.

^{iv} King, Keohane and Verba (1994) provide an excellent account of the use of qualitative methods for investigating causal hypotheses.

Improving the evidence base for energy policy

When EBPP is extended to other areas of policy and practice where (quasi-) experimental studies are less feasible or appropriate, the consensus over the relative weight to give to different research methodologies can break down.^v Table 1 provides a stylised distinction between those areas of policy and practice where a consensus broadly exists and those where it does not. In practice, the situation may be better represented as a spectrum of possibilities, with disputes arising even within ‘core’ areas of EBPP such as healthcare. Furthermore it is possible to adopt widely different methodological approaches to very similar research questions.^{vi}

Table 1 Contrasting attitudes to EBPP in different areas of policy and practice

Issue	‘Consensus’ area of policy research	‘Disputed’ area of policy research
Outcomes of policies and interventions	Simple	Complex
Objectives of policies and interventions	Simple, reinforcing and consensual	Multiple, competing and contested
Dominant research methodology	Experimental and quasi experimental	Qualitative methods
Degree of consensus regarding appropriate research methodology	High	Low
Importance of context in determining outcomes	Low	High
Emphasis given to evaluating policies and interventions	High	Low
Typical objective of such evaluations	Whether the policy/intervention works	Why, how and in what circumstances the policy/intervention works

These distinctions are important when assessing the potential role of EBPP within energy policy. For example, where on this spectrum does energy policy, or different approaches to energy policy, lie? Table 2 compares the methodological approaches in energy policy research with those taken in other policy areas (Davies, Nutley *et al.*, 2000),.

^v “...In some areas (most notably healthcare) the need for evidence and the nature of convincing evidence is a given. In other areas (most strikingly, social care), the very nature of evidence is hotly disputed and there is strong resistance to assigning privileged status to one research method over another. Such divergent attitudes arise from deep-rooted ontological and epistemological assumptions. Furthermore, where post-modern perspectives are prevalent there is a general distrust of any notion of objective evidence.” (Davies, Nutley *et al.*, 1999)

^{vi} Compare for example sociological approaches to household energy consumption (Hand, Southerton *et al.*, 2003) to econometric estimates of household production functions (Willett and Naghsipour, 1987).

Table 2 Methodological preferences and debates in different policy areas

Policy area	Methodological preferences and debates
Healthcare	Gold standard of randomised controlled trials with additional methodological safeguards. Growing interest in qualitative methods to give a complementary view.
Education	Competing rather than complementary methods with methodological disputes between different groups. Econometric analysis of large data sets, but little experimental research.
Criminal justice	General acceptance of experimental or quasi-experimental methods.
Social care	Preference for qualitative methodologies. Quantification and experimentation often viewed with suspicion and hostility.
Welfare policy	Eclectic use of methods to provide complementary insights. Some longitudinal study but almost no experimentation.
Housing	Predominant use of qualitative and quantitative survey methods. Use of econometric methods for forecasting housing needs. Recent emergence of more multi-disciplinary approaches.
Urban policy	Difficulties in relating particular outcomes to particular interventions and in identifying externalities. Diverse methods employed, but particular reliance on case studies. Little or no experimentation
Transport	Multidisciplinary. Policy research frequently rooted in economic modelling and forecasting.
<i>Energy policy</i>	<i>Multidisciplinary, but with a strong bias towards econometric analysis of secondary data and economic modelling. Policy evaluation relatively weak in many areas.</i>

Source: Adapted from Davies *et al* (2000). Text on energy policy added.

EBPP practitioners have identified a number of problems with the existing systems for creating, synthesising, disseminating and utilising research evidence and have provided a number of suggestions on how these may be improved. But the greatest weakness lies in the inadequate accumulation and synthesis of research results and the poor quality of traditional literature reviews. This is a dominant theme within EBPP and the primary rationale for developing systematic reviews of the available evidence base.

3. The methodology of systematic reviews

There are several reasons why systematic reviews have become so central a feature of EBPP. First, experience in the medical field and elsewhere suggests that policy and practice are often based on inadequate evidence. For example, the UK Department for Health (1991) found that only around 15% of the clinical interventions used in the National Health Service were based upon unequivocal scientific evidence and many commonly used interventions did not improve the health of patients (some, indeed, appeared to make their condition worse). In addition, interventions that were well supported by evidence were often neglected or ignored.^{vii} Second, the increasing volume of research findings makes it difficult for policy makers and practitioners to keep abreast of current understanding, creating a need for more effective synthesis of research results. Third, a combination of the complexity of the relevant issues, the variable quality of research evidence and the methodological and other biases of individual researchers, leads to conflicting recommendations by different authors and

^{vii} The same problem is found in other areas. For example, prison tour programmes for preventing juvenile delinquency have enjoyed a great deal of popularity, despite only anecdotal evidence of their success. A systematic review of the subject found that the programmes had the opposite effect from that intended – they made crime *more* likely (Petrosino, Turpin-Petrosino *et al.*, 2000).

Improving the evidence base for energy policy

corresponding uncertainty over whom to trust. This problem can be exacerbated by the selective use of evidence by powerful interest groups and by the partial and unbalanced treatment of research results by the media. Finally, the traditional form of research synthesis – termed the *narrative review* by EBPP practitioners – often fails to address these limitations, with the result that different reviewers can reach different conclusion from the same research base.^{viii} More specifically, the criticisms levelled at traditional narrative reviews include (Petticrew and Roberts, 2005):

- Poor specification of the review topic, leading to excessively wide-ranging discussion and inconclusive results.
- Selective and opportunistic use of evidence, leading to selection bias and the neglect of relevant studies.
- Inadequate specification of the criteria for including or excluding studies from a review.
- Limited attention to methodological quality, leading to a lack of discrimination between sound and unsound studies.
- Lack of transparency, encouraging subjectivity and bias in the reporting of results.

Both research practitioners and commentators on EBPP have disputed these criticisms (Hammersley, 2001), but there is evidence to suggest that they are valid. For example, Oliver *et al* (1999) compared six different narrative reviews of accident prevention for older people, covering a total of 137 studies. Of these, only 33 studies were common to at least two reviews, only two were common to all six reviews and only one study was treated consistently by all six. Similarly, Oakley and Fullerton (1995) found 70 studies of smoking prevention programmes for young people, of which only 27 had been captured by two previous narrative reviews. Of these, only three were common to both reviews and the differences in coverage led the two reviews to reach different conclusions on the relative effectiveness of different prevention programmes.

Systematic reviews of existing evidence seek to address each of the above criticisms. Ideally, the systematic review will reach an authoritative conclusion regarding a specific research question through a comprehensive synthesis of currently available primary research. But where this is not possible, the review will seek to explain the reasons for the differences between studies and to identify priority areas for further research.

Systematic reviews use explicit and transparent methodologies that are replicable and updateable. They involve:

- A clear specification of the research question(s) to be addressed.
- Systematic and exhaustive searching of the available literature.
- Applying explicit criteria for the inclusion or exclusion of studies.
- Appraising the quality of the included studies using transparent and standardised criteria.
- Summarising and synthesising the results in an ‘objective’ manner.
- Disseminating the results effectively to the appropriate audience.
- Updating the results of the review at intervals, when new research becomes available.

^{viii} Oakley (2002) is particularly disparaging: ‘...Most traditional literature reviews are discursive rampages through selected bits of literature that the researcher happens to know about or can easily reach on his or her bookshelves at the time’.

Improving the evidence base for energy policy

Prior to conducting a systematic review, the review team will publish their intended approach to each of these issues in a *Protocol*. This will describe the scope of the review, the strategy for searching the evidence base, the criteria for the inclusion and exclusion of evidence, the basis for assessing the quality of such evidence and the procedure for synthesising the results. The Protocol will be open to comment by relevant stakeholders, with the aim of improving transparency and credibility. Organisations such as the Campbell Collaboration have published guidelines for producing such Protocols, together with libraries of examples. The rigorous and comprehensive nature of a systematic review leads to correspondingly greater resource requirements than for a comparable narrative review, which may create difficulties for research funders.^{ix}

The differences between systematic reviews and traditional narrative reviews are summarised in Table 3. The following sections describe each stage of a systematic review in more detail.

Table 3 Claimed differences between systematic and narrative reviews

Stage	Good-quality systematic reviews	Traditional narrative reviews
Deciding on review questions	Start with clear question to be answered and/or hypotheses to be tested	May start with a clear question to be answered, but more often involve general discussion of subject with no stated hypotheses.
Searching for relevant studies	Strive to locate all relevant published and unpublished studies to limit selection bias	Do not usually attempt to locate all the relevant literature
Deciding which studies to include or excluded	Include explicit description of what types of studies are to be included to limit selection bias	Usually do not describe why some studies are included and others excluded
Assessing study quality	Examine in systematic manner the methods used and investigate potential biases and sources of heterogeneity between study results	Often do not consider differences in study methods or quality
Synthesising results	Base conclusions on the studies which are considered to be most methodologically sound	Often do not differentiate between methodologically sound and unsound studies
Replicating and updating	Use protocols and explicit criteria to ensure that others would reach the same conclusions if they adopted the same methods, and so the results may easily be updated.	Use methodologies and criteria that lack transparency, leaving the interpretation of results open to bias.

Source: Pettricrew (2001)

Choosing the review questions

In contrast to traditional literature reviews, systematic reviews usually employ a sharply defined research question, or a hypothesis that is suitable for test through empirical research. The UK Cabinet Office (2005) recommends that these questions specify the mechanisms, interventions or policies in question; the population and/or subgroups in question; the outcomes that are of interest; and the context in which the question is set. For example, rather than reviewing the literature on ‘barriers to energy efficiency’, a systematic review should seek to answer a more specific question such as:

^{ix} The EPPI Centre, for example, recommends as a minimum two researchers at 50% full-time equivalent for nine months, together with additional resources for information scientists.

‘What is the evidence that the inability to appropriate the cost savings from energy efficiency investment (mechanism) leads households in rental accommodation (population) to use energy less efficiently than owner occupiers (outcome)?’

Review questions are normally refined to make them more specific and elaborated by specifying the objectives more clearly.^x In practice, systematic reviews tend to focus primarily on the effects of individual policies or interventions, as opposed to other types of mechanism, and to prioritise the use of (quasi-) experimental evidence, as opposed to other types of evidence. An example of a quasi-experimental study would be a before and after comparison of energy consumption by participants in a demand-side management (DSM) scheme, with or without a control group of non-participants (Meyer, 1995).

Searching the literature

Systematic reviews aim to be comprehensive in their coverage of the available literature, including both peer reviewed academic papers and ‘grey’ literature such as PhD dissertations, conference papers and consultants’ reports. However, while non-peer reviewed literature is generally admissible, the absence of peer review may be one factor taken into account when assessing the quality of the studies. Similarly, while systematic reviews should not be restricted to English language sources, resource constraints often necessitate this in practice.

The Protocol specifies the databases, bibliographies, contacts and other sources that are to be employed; the years to be covered; the search strategies to be used (e.g. keywords); and the mechanisms available for retrieving less accessible documents. In practice, the relative importance of different information sources and search strategies will vary between different fields.^{xi} Typically, the literature search is resource intensive and requires the support of information specialists. Also, many more studies are normally identified than are finally used in the review. For example, in reviewing the evidence on the effectiveness of policy measures to prevent homeowners defaulting on their mortgage, Wallace *et al* (2004) initially identified a total of 1832 references. Of these only 766 were classified as generally relevant, 49 met the inclusion criteria and only 22 eventually passed the quality threshold.

Including and excluding studies

In parallel with defining specific questions, systematic reviews establish criteria for including and excluding studies from the review. For example, in the mortgage default study cited above, the inclusion criteria were that the study should: a) address at least one of the four types of policy intervention defined in the protocol; b) include assessment of policy effectiveness; and c) be based upon empirical research (Wallace, Croucher *et al.*, 2004).

Applying criteria such as these can lead to the exclusion of a large number of the identified studies. For example, the mortgage study excluded 97% of the identified references, largely because they did not include empirical research. The exclusion of evidence on these grounds has been criticised by commentators who consider that expert opinion, theoretical models and

^x For example, in a systematic review of the ‘effects of neighbourhood watch schemes on preventing crime’, the following objectives were established: a) to operationalise the inputs (e.g. the policies qualifying as neighbourhood watch) and the outcomes (e.g. the relevant types and measures of crime); b) to identify studies that evaluate the effect of the qualifying schemes on the relevant outcomes; c) to identify a list of studies that meet a minimum criteria of scientific rigour; d) to obtain a comparable measure of effect size in the selected most rigorous studies; and e) To arrive at a conclusion about the effectiveness of neighbourhood watch schemes (Bennett, Farrington *et al.*, 2004).

^{xi} For example, indexed and comprehensive electronic databases are more established in healthcare than in most areas of social science. Keyword searching is also more difficult in the social sciences, owing to the relative imprecision and inconsistency in the use of language.

other types of literature may be very relevant to particular research questions. In practice, many systematic reviews will use such studies as ‘conceptual background’ and will summarise the conclusions of these studies separately from the main text.

As with defining the research question, the choice of inclusion criteria involves a difficult compromise between comprehensiveness and quality control. If strict criteria are used, there is a risk that the majority of relevant studies will be excluded, but if loose criteria are used, synthesis will become more difficult.

Extracting information

Systematic reviews extract and store information from the included studies in a standardised way. Typical headings include the nature of mechanisms studied; the research methods used; the choice of population and sample; the methods for controlling intervening variables; the outcomes measured or observed; and the size of the effects. Several organisations provide standardised tools and databases for this purpose.

Most of the standard frameworks are designed for quantitative studies, often with the aim of conducting a subsequent **meta-analysis** of results (see below). Similarly, most of the standard frameworks are designed for studies that evaluate the outcomes of particular policies or interventions and are poorly suited to other types of evidence.

The extraction process is normally conducted prior to the quality assessment of individual studies, which means that much of the extracted information may be discarded when drawing the final conclusions. However, the review database can provide a valuable reference source for the status of research in a particular area, and can make it easier for researchers to compare the results of different studies and to identify gaps.

Appraising quality

Explicit appraisal of the quality of individual studies is perhaps the defining feature of a systematic review, but also the most controversial. Most reviews use a weighted or un-weighted checklist of requirements against which each study is scored. Only studies that exceed a certain quality threshold are retained for use in the final synthesis.

A range of quality criteria have been developed, with most having a bias towards (quasi-) experimental studies. Three standard concepts that dominate are:

- *Internal Validity*: The extent to which the study shows a cause-effect relationship between the independent and dependent variables.
- *Construct validity*: The adequacy of the operational definition and measurement of the theoretical constructs.
- *External validity*: The ‘generalisability’ of the proposed causal relationships among different actors, places and times.

These may be refined into very specific criteria if a single type of quantitative evidence dominates,^{xii} but may need to be stated in more general terms if multiple sources of evidence

^{xii} As an illustration, a review of quasi-experimental studies in criminology used the following classification scheme (in ascending order of quality): 1) correlation between a prevention programme and a measure of crime at one point in time; 2) measures of crime before and after the programme, with no comparable control condition; 3) measures of crime before and after the programme in experimental and comparable control conditions; 4) measures of crime before and after the

Improving the evidence base for energy policy

(including qualitative studies) are to be included. Table 4 lists the criteria used to appraise the quality of both quantitative and qualitative studies in the systematic review of ‘mortgage safety nets’ cited earlier (Wallace, Croucher et al., 2004). This list demonstrates that quality appraisal is necessarily a matter of judgement (e.g. just when can a description of context be considered ‘adequate’?). However, the importance of subjectivity may vary with the type of study and the transparency of the appraisal process can be valuable in itself. Most systematic reviews seek to minimise subjectivity by having two or more researchers conduct quality appraisals and compare results.

Table 4 Criteria for appraising the quality of both quantitative and qualitative studies

Area	Issue	Essential/ Desirable
Question	Is the research question clear?	Essential
Theoretical perspective	Is the theoretical perspective of the author(s) explicit and has this influenced the study design, methods or findings	Desirable
Study design	Is the study design appropriate to answer the question?	Essential
Context	Is the context or setting adequately described?	Desirable
Sampling	<i>Qualitative</i> : is the sample adequate to explore the range of subjects and settings and has it been drawn from an appropriate population? <i>Quantitative</i> : is the sample size adequate for the analysis used and has it been drawn from appropriate population?	Essential
Data collection	Was the data collection adequately described and rigorously conducted to ensure confidence in the findings?	Essential
Data analysis	Was there evidence that the data analysis was rigorously conducted to ensure confidence in the findings?	Essential
Reflexivity	Are the findings substantiated by the data and has consideration been given to any limitations of the methods or data that may have affected the results?	Desirable
Generalisability	Do any claims to generalisability follow logically, theoretically and statistically from the data?	Desirable
Ethics	Have ethical issues be addressed and confidentiality respected?	Desirable

Source: Croucher, Quilgars et al. (2003)

The need for dedicated criteria to appraise qualitative research has become more pressing as EBPP techniques are extended to a wider range of policy areas.^{xiii} In response, numerous review teams have developed checklists of criteria, often with considerable differences in approach.^{xiv} In an attempt to introduce some standardisation, the Cabinet Office Strategy Unit has published a comprehensive review of criteria for assessing qualitative research and has proposed a synthesising framework involving 18 appraisal questions and 64 indicators (Spencer, Ritchie et al., 2003).

More recent approaches have sought to extend appraisal beyond issues of methodological quality to include the quality of reporting and relevance to important policy questions (Boaz and Ashby, 2003). This recognises that excessive concentration on methodological quality may create the dual risk of including rigorous studies that are of little value to stakeholders and excluding studies that are highly relevant but have a weaker methodology. But since this

programme in multiple experimental and control units, controlling for other variables that influence crime. 5) random assignment of programme and control conditions to units.

^{xiii} This need has also become apparent within ‘core’ EBPP applications such as healthcare. For example, a systematic review of the influences on the take-up of childhood immunisations found that several key determinants were missed if only quantitative studies were included (Roberts, Dixon-Woods et al., 2002).

^{xiv} For example, Oakley (2002) compared the criteria proposed by four different review teams: out of a total of 46 criteria, 28 occurred in only one of the lists, 10 in two and six in three, with only two criteria being common to all four lists.

departs from one of the core tenets of systematic reviews - giving priority to methodologically rigorous studies - it has encountered resistance.

Synthesising results

The final stage of a review (excluding dissemination and updating) is the synthesis of results from the subset of studies that pass the quality appraisal. If these are quantitative studies, the synthesis can be conducted in a rigorous way, using a group of statistical techniques known as *meta-analysis* (Cooper and Hedges, 1994). The objective may either be to obtain a more accurate estimate of the existence and size of the relevant effect, or to explain the variations in results between studies. By combining the results of several studies, the overall sample size may be increased and statistical tests may have a greater probability of detecting the relevant effects (i.e. greater power).

Meta-analysis would appear to be particularly suitable for the synthesis of certain types of econometric data, such as estimates of price or income elasticities. To date, however, the level of take-up within the economics profession appears limited (Florax, de Groot *et al.*, 2002). Meta-analysis is much less suitable when there are significant differences in the mechanisms or policies under investigation, the outcomes measured, or the context in which the studies took place. In these cases, a narrative synthesis likely to form an important part of any review, but the EBPP literature provides much less guidance on how this should be achieved. Where quantitative and qualitative research is combined there is a tendency to prioritise the first and use the second to aid the interpretation of the quantitative findings.

4. The limitations of systematic reviews

Systematic reviews have been criticised by researchers and practitioners in a number of fields (Hammersley, 2001). Some of these criticisms are misconceived and have been effectively countered by EBPP advocates (see Box 2), while others have their roots in long-running disputes within the philosophy of science. These criticisms have been applied to the original application of systematic reviews - experimental studies of highly specific medical interventions - but become more vigorous when the approach is extended to complex policy issues. Research in these areas is frequently qualitative and/or characterised by disputes between competing disciplines.

Box 2 'Myths' about systematic reviews

- *Systematic reviews are the same as ordinary reviews, only bigger:* While they are more comprehensive, features such as specificity, transparency and quality appraisal also make them qualitatively different.
- *Systematic reviews include only randomised controlled trials:* A variety of quantitative and qualitative methodologies have been included within SRs.
- *Systematic reviews are of no relevance to the real world:* Systematic reviews have been used to examine a range of contemporary and contentious policy issues such as domestic violence and child abuse.
- *Systematic reviews necessarily involve statistical synthesis:* Many systematic reviews do not use statistical methods.
- *Systematic reviews are a substitute for individual studies:* Systematic reviews do not necessarily provide definitive answers and often identify the need for additional primary research.

Source: Petticrew (2001)

In this context, it is unsurprising that the assumptions of systematic reviews should encounter difficulties. What is less clear is whether, as a result, the approach has nothing to offer policymakers and researchers, or whether it can be usefully adapted to address interesting and relevant policy questions. This section highlights three interrelated ‘biases’ in the systematic review methodology, namely:

- bias in the selection of question;
- bias in the selection and appraisal of evidence; and
- bias in the synthesis of results.

Bias in the selection of question

Systematic reviews require clear, specific and answerable questions that are tightly and narrowly defined. At the same time, the questions should be relevant to current debates and of sufficient significance to justify the resources spent. But there is a tension between these objectives: narrow questions may be more answerable but of less interest to policymakers and practitioners (especially given the resource requirements of a systematic review), while broad questions may be more interesting but less amenable to the systematic approach.

Systematic reviews commonly address ‘micro’ questions regarding the technical efficiency and effectiveness of particular policies or practices, but they rarely address ‘macro’ policy questions, or those with a substantial normative content. Since such questions are relevant to all areas of policy, the bias results more from the constraints of the methodology, than from the characteristics of the policy sectors to which it has been applied.

Most systematic reviews formulate a question of the form: ‘What is the effect of policy/intervention X in population Y on outcome(s) Z?’ This focus on ‘what works’ reflects both the medical origins of EEBP and its strong links with policy evaluation. But ‘what works’ is neither the only question of interest to policymakers or necessarily the most important question. For example, other relevant questions may include: What are the recent trends in X? What are the causes of X? What are the risks of X? What are the likely costs and benefits of X? What might happen if X is done? (Solesbury, 2001, p. 8).

A useful distinction here is between *descriptive* questions – such as: ‘what are the recent trends in UK industrial fuel consumption?’ and *causal* questions – such as: ‘what are the reasons for those trends? Both provide useful information for policymaking and both involve theoretical assumptions – for example, in defining measurement categories or in choosing between decomposition methodologies. But historically, systematic reviews have focused almost exclusively on the latter.

A possible way of subdividing causal questions is to distinguish between *historical* questions, which enquire into the reasons for present or historical states of the world, and *forecasting* questions, which estimate future states of the world under given assumptions. Both are relevant to policymaking, but systematic reviews have focused almost exclusively on the former.

The sources of evidence for causal questions include both *theoretical* studies that postulate, model or explain relationships between variables and *empirical* studies that calibrate or test those frameworks using historical data. Both are relevant to causal questions, although individual studies may be either largely theoretical or largely empirical, as well as either

largely quantitative or largely qualitative. But systematic reviews have focused primarily on empirical studies, and particularly those that use quantitative methodologies

These distinctions (summarised in Table 5) are relevant to the potential application of systematic reviews to energy policy. While systematic reviews have traditionally focused on historical questions, energy policy researchers are preoccupied with forecasting questions such as the technical and economic potential of particular technologies. Similarly, while systematic reviews have traditionally focused on empirical questions, such as evaluating the success (under some measure) of particular policies, energy policy researchers have spent a great deal of time on theoretical questions. The distinction is relative rather than absolute; since all empirical questions assume a particular theoretical framework and individual studies may test, develop or modify these frameworks as appropriate. Nevertheless, systematic reviews have tended to assume that there is broad consensus on the appropriate framework and hence provide relatively little guidance on how to adjudicate between competing theoretical approaches.

Energy policy research does share common ground with systematic reviews in that priority tends to be given to quantitative methodologies. However, systematic reviews have a bias towards (quasi-) experimental studies while energy research has a bias towards various forms of economic modelling. These models are based in varying degrees upon real-world data, but also embody a large number of theoretical assumptions. While a systematic comparison of modelling results has been undertaken (Barker, Koehler et al., 2002), there appears to be little precedent for this within the systematic review literature.

Table 5 Classifying types of question, study and methodology

	Dominant in systematic reviews	Less common in systematic reviews
Purpose of question	Causal	Descriptive
Focus of question	Historical	Forecasting
Nature of study	Empirical	Theoretical
Dominant methodology	Quantitative	Qualitative

Bias in the selection and appraisal of evidence

Systematic reviews were originally applied to ‘what works’ questions within medicine. These are best addressed through experimental studies, utilising explicit and replicable procedures that allow for physical or statistical control of intervening variables, together with the generalisation of results. This notion of a ‘gold standard’ methodology has influenced the application of systematic reviews in all areas, even where experimental studies are less feasible. As a result, this approach has been criticised for its allegedly ‘positivist’ assumptions, including: first, that society should be studied in the same way as the natural world; second, that causality may be established by identifying empirical regularities among sequences of events; third, that some research methodologies are better than others in all instances; and fourth, that subjectivity is a source of bias that should be minimised (Hammersley, 2001).

Each of these assumptions has been the subject of sustained criticism within the philosophy of science for many years. Some practitioners reject all of them and argue that the appropriate aim of social science is to elucidate meanings. But most take a compromise position, emphasising the differences between natural and social science but still claiming that causality has meaning in social systems and can be identified in particular instances.

Improving the evidence base for energy policy

However, since the social and natural worlds are fundamentally different, there are likely to be comparable differences in the methodologies that can be employed and the status of the knowledge claims that can be made (Sayer, 1992).

In natural science, causality is normally established through identifying regularities in the outcomes of repeated experiments. But doing the same for public policies or complex social interventions can be problematic. Regularities are achievable in natural science because experiments establish 'closed' systems where the mechanism possessing the causal power is stable, and where the external conditions in which the mechanism is situated are held constant (Bhaskar, 1975). These conditions are practically impossible to achieve in social situations because: first, the relevant mechanisms involve the actions of people who can reason, learn and change their behaviour; and second, there are host of complex contextual influences that are difficult to control. As a result, the relationship between particular causes and effects is like to vary across time and space. The same causal mechanism may produce different outcomes according to context, the same outcome may be produced by different causal mechanisms, and the operation of different mechanisms may not be stable since they depend upon the understanding and behaviour of individuals. These differences have led some authors to argue that the scope for 'generalisation' within social science is limited and the best that can be hoped for is the identification of partial regularities that hold for only a limited period of time (Lawson, 1997). Hence, the fact that a particular relationship between cause and effect does not hold in a particular instance may not be a reason to reject the underlying theory, because a host of offsetting contextual influences may be at work.

This stance leads to a preference for qualitative rather than quantitative research methods, and an emphasis on the theoretical content of explanations as much as the statistical relationships between observed variables (Sayer, 1992). Rival theories are then judged not solely in terms of their empirical success but also in terms of their 'explanatory adequacy' (Lawson, 1997).^{xv} However, this preference has been challenged by Pratschke (2003) and others who argue that statistical approaches such as econometrics are entirely compatible with this more nuanced understanding of the nature of causality in social systems (Bache, 2003, ; Hodgson, 2004). This is largely because there are reasons to expect *relatively* enduring regularities in some social systems, and methods such as econometrics can reveal such regularities while controlling as far as possible for intervening variables:

“... although the social world is potentially, or logically, a purely open system, in practice it is characterised by 'quasi-closures' which may manifest themselves in stable patterns of events....Faced with the challenge to make decisions in complex situations, agents are inclined to develop habits, conventions and routines. Such habitual and routine behaviour will be reflected in social institutions which may prove to be stable over periods of time.” (Bache, 2003, p. 14)

This methodological debate is ongoing. But one possible conclusion is that the research methods should be *appropriate* to the subject matter under investigation - with no one method being necessarily superior to any other. While some EBPP authors accept this, there is still a tendency to prioritise (quasi-) experimental studies within a 'hierarchy' of methodologies (Hammersley, 2001). Qualitative studies are often given a 'supporting role', either providing mere 'conceptual background' or being summarised separately from the main meta-analysis. But the contemporary philosophy of science provides little justification for this and it can present practical difficulties when systematic reviews are applied to questions where qualitative research (quite appropriately) dominates. As a result, Boaz and Ashby (2003) have

^{xv} The philosophical approach being sketched here was first developed by Bhaskar (1975) and is termed critical realism (Archer, Bhaskar *et al.*, 1999). 'Explanatory adequacy' is a core concept in critical realism, but is poorly defined. For a discussion, see (Peacock, 2000).

advocated the replacement of methodological hierarchies with the notion of *fitness for purpose*.

A related aspect of the ‘anti-positivist’ critique is the necessity of judgement, both within scientific research and within the systematic review process itself. As argued by Hammersly (2001), the systematic review methodology has tended to assume that subjectivity is a source of bias that should be minimised by transparency and procedural rules. But the ‘post-positivist’ perspective argues that judgement is unavoidable and need not necessarily be a source of bias. Instead, informed judgement can involve skilled and knowledgeable assessment of what is likely to be true and hence can help to identify mistakes and weed out incorrect and inconsistent assumptions. As a result, both the attempts to minimise judgement within the systematic review process, and the priority given to research evidence as compared to professional knowledge and experience, may be misguided (Hammersley, 2005). As such, this critique questions the rationale of the ‘evidence-based’ process itself.

Bias in the synthesis of results

Meta-analysis commonly seeks to synthesise the results of various studies by ‘adding’ them together and using the larger effective sample size to obtain more accurate estimates of the size of the relevant effect. This ‘additive’ idea also influences the traditional approach to narrative synthesis. But such an approach is only possible if each of the studies address the same specific issue and investigate it in a sufficiently similar way that the findings can be aggregated (Hammersley, 2001). For complex policies and mechanisms operating within heterogeneous social contexts, this may not be viable.

As Hammersly (2001) has argued, the ‘additive’ approach represents a rather impoverished understanding of what synthesis means. Synthesis is more usually understood as combining a set of parts into a whole, which may be greater than the sum of parts. This may involve conceptual development that goes beyond the original studies and is likely to require creative interpretation, rather than simply following a procedure. For example, a study on the effect of energy taxes on household energy use could seek to combine econometric estimates of energy price elasticities, survey data on perceived barriers to technology adoption and behavioural studies of decision-making by individual households. These studies may complement one another, in that they can reinforce or challenge each other's assumptions and conclusions (Hammersley, 2005).

Hence, producing a good literature review is not simply a matter of ‘summing’ data, but instead involves judging the validity of different theoretical claims and empirical findings and thinking about how these relate to one another (Hammersley, 2005). An implication of this is that useful syntheses may not require an exhaustive search of the relevant literature, since an attempt to do so may simply lead to diminishing returns. Instead, a more valuable approach may be to use studies in one area to provide insights into problems and issues in another.

5. Is energy different?

Earlier sections have highlighted several differences between energy policy and those areas where EBPP is most established. These may usefully be summarised as follows:

- *Experimental versus non-experimental*: EBPP has proved most successful in those areas of policy and practice where there is a strong tradition of (quasi-) experimental research.

Improving the evidence base for energy policy

But this is less well represented in energy research, which tends to rely more on econometric analysis of secondary data and economic modelling.

- *Micro versus macro*: EBPP has proved most successful for policies and interventions that are focused at the individual or community level – such as in education and social work. These are also the areas that are most suitable for quasi-experimental research methodologies. But much of energy policy is focused at the sector level and has economy-wide implications.
- *Means versus ends*: EBPP has proved most successful in those areas of policy and practice where there is consensus on a small number of objectives and where the primary topic of debate is the means for achieving those objectives. The contribution of evidence in this context is primarily to establish ‘what works’. But energy policy is characterised by multiple and often competing objectives whose relative importance varies over time and is frequently the topic of fierce debate.
- *Service provider versus regulator*: EBPP has proved most successful in those areas of policy where the government acts as a service provider - such as healthcare and education. But the government's main role within energy policy is as a regulator and facilitator of competitive markets (Majone, 1997).^{xvi}
- *Practice versus policy*: EBPP has proved most successful in influencing the practice of doctors, teachers, social workers and other professionals who need to integrate evidence into their everyday decision-making. But with energy policy, politicians, civil servants and regulators focus primarily upon strategic decisions that will shape policy for a considerable period of time.
- *Evaluation versus forecasting*: EBPP has proved most successful in areas where there is a strong tradition of policy evaluation. But energy policy research has placed greater emphasis on economic forecasting than policy evaluation, with the result that the evidence base is weaker.
- *Technical versus politicised*: EBPP has proved most successful in specialised areas of policy or practice, where there is relatively little politicisation - for example, the choice of therapy for a particular disease. But the key questions within energy policy have substantial economic and political consequences and are subject to influence from powerful interest groups. Here, evidence tends to be used to support particular values, interests and policy prescriptions, rather than to underpin decision-making in a more neutral way.
- *Confirmation versus innovation*: EBPP has proved most successful for synthesising evidence from numerous studies that use very similar methodologies. But the funders of energy research increasingly seek empirical, conceptual and methodological innovation, leading to a diversity of approaches that can make the meta-analysis of results problematic.

Each of the above distinctions is an oversimplification and numerous exceptions can be found. For example: systematic reviews have been applied to ‘macro’ policy issues such as inequality and violent crime (Hsieh and Pugh, 1993); to non-service provider areas such as tax policy (Blundell and Walker, 2000); to non-evaluation questions such as the impact of race on sentencing (Sweeney and Haney, 1992); and to highly politicised issues such as GM crops (Levitt, 2003). But while precedents exist in each of these areas, they are often the exception rather than the rule. Similarly, while systematic reviews have been applied to policy areas that have similarities with energy (e.g. transport), the experience to date in these areas is

^{xvi} The author is indebted to Jim Skea, Research Director of the UKERC, for this insight.

very limited (Kremers, Nijkamp et al., 1999). This suggests that systematic review techniques may need to be modified if they are to be successfully applied to energy policy.

An example: the rebound effect

The difficulties of applying systematic reviews to energy policy questions may briefly be illustrated with an example. A topic that is attracting increasing attention is the magnitude of the ‘rebound effect’ from energy efficiency improvements. Since improved energy efficiency reduces the effective price of an energy service, there may be an increase in consumption of that service that could offset the energy savings achieved. The issue in question is the size of this effect for a particular energy service, such as household heating or passenger transport.

At first sight, the application of systematic review techniques appears sensible. A large number of empirical studies of the rebound effect are available for both household heating and passenger transport and many of these use quasi-experimental techniques that are suitable for appraisal using standard EBPP criteria (Greening, Greene *et al.*, 2000). But a closer look reveals that this evidence base is extremely diverse, both empirically and methodologically. The relevant econometric studies use time series, cross-sectional and panel data from a variety of time periods and geographical areas and estimate parameters using a range of functional forms and controlling for a range of intervening variables (Small and Van Dender, 2005). Differences in results owe much to differences in functional specification or empirical application (e.g. time periods before or after oil price shocks), making it difficult to use meta-analysis to estimate the ‘true’ effect (Greening, Greene *et al.*, 2000). While these differences could potentially be explored through meta-regression techniques (Espey, 1998), this would need to accommodate both quasi-experimental studies and the larger number of *indirect* estimates of the rebound effect that are based upon price elasticities.^{xvii} The latter approach rests upon a number of questionable assumptions that may limit both the accuracy and comparability of different studies (Sorrell and Dimitropoulos, 2005). Indeed, the most useful contribution of a review in this area may be to clarify these underlying theoretical issues.

A systematic review would only be possible for the small number of energy services that have been the subject of significant research on rebound effects. Moreover, the precise measurement of the *direct* rebound effect for a particular energy service may not be the most interesting policy question. This is because the ‘full’ rebound effect also includes:

- *Indirect effects*: The lower effective price of the energy service can lead to changes in the demand for other goods, services and factors of production that also require energy for their provision. For example, the cost savings obtained from a more efficient central heating system may be put towards an overseas holiday.
- *Economy wide effects*: A fall in the real price of energy services may reduce the price of intermediate and final goods throughout the economy, leading to a series of price and quantity adjustments, with energy-intensive goods and sectors gaining at the expense of less energy-intensive ones. Energy efficiency improvements may also reduce energy prices and increase economic growth, which could further increase energy consumption. (Greening, Greene *et al.*, 2000).

^{xvii} This makes use of the ‘Khazzoom’ equation: $\eta_\epsilon(E) = -\eta_{P_\epsilon}(S) - 1$, where $\eta_\epsilon(E)$ is the elasticity of energy demand with respect to energy efficiency alone and $\eta_{P_\epsilon}(S)$ is the elasticity of energy service demand with respect to the unit price of energy services. Frequently, energy demand (E) is substituted for energy service demand S , and/or the unit price of energy commodities (P_ϵ) is substituted for the unit price of energy services (Sorrell and Dimitropoulos, 2005).

Improving the evidence base for energy policy

The controversy over the rebound effect relates to the aggregate impact of all these mechanisms. Brookes (2000) and Saunders (2000), for example, argue that improved energy efficiency may actually lead to *higher* energy consumption overall ('backfire').

Systematic review techniques may have little to contribute to this broader, economy-wide question. In addition to the quasi-experimental and price elasticity studies, the relevant evidence base includes: life-cycle and input-output analysis of the energy content of (Alfredsson, 2004); decomposition analysis of historical trends in energy consumption (Schipper and Grubb, 2000); general equilibrium modelling of the macro-economy (Grepperud and Rasmussen, 2004); econometric estimates of the elasticity of substitution between energy and capital (Berndt and Award, 1979); and neoclassical growth theory (Saunders, 2000). Several of these categories have not previously appeared within EBPP, so there is little precedent for comparative evaluation of their methodological quality. Moreover, the overall evidence base is too large in terms of the number of studies, too heterogeneous in terms of approach and empirical application and (most importantly) too embedded in complex and contested theoretical frameworks to be an appropriate subject for a systematic review (Table 6). The central importance of theoretical issues means that a useful review would need to do far more than 'combine' different studies to estimate the 'true' size of the effect.

Table 6 An overview of the evidence base for the rebound effect

Methodological category	Size of evidence base	Internal diversity	Theoretical content	Suitability for systematic review
Direct measurement	Large	Large	Medium	High
Price elasticities	Very large	Very large	High	Medium
LCA analysis	Small	Small	Medium	Low
Decomposition analysis	Very large	Medium	Medium	Low
General equilibrium modelling	Small	Medium	High	Low
Neo-classical growth theory	Small	Small	High	Low
Elasticities of substitution	Large	Large	High	Medium

Source: Sorrell and Dimitropoulos (2005)

This brief review suggests that systematic reviews face methodological difficulties when applied to highly specific questions regarding direct rebound effects, and are entirely inappropriate for a review of the 'full' rebound effect. While the rebound effect may be a particularly difficult topic, it is likely that other policy relevant questions in the energy field have similar characteristics. Similarly, while systematic reviews may be inappropriate for such multidimensional questions, these are also the questions most likely to attract research funding. Hence, the more widespread use of systematic reviews in energy research may require a change in funding priorities.

6. Summary

Energy policy research is not immune from the problems that EBPP set out to address. These include: conflict and confusion over key issues; over-reliance on individual studies; inadequate accumulation and synthesis of research results; and wide-ranging and inconclusive literature reviews that pay insufficient attention to methodological quality (Gross, 2005). To

the extent that systematic reviews have a track record of successfully overcoming these difficulties, there ought to be scope for applying this approach to energy policy questions.

However, systematic reviews have a number of important weaknesses. These include: the narrow range of questions to which they have been applied; the bias towards quantitative research methodologies; the difficulties in addressing complex problems and policies; and the ‘additive’ approach to synthesis that neglects the complementary nature of different studies. Furthermore, there appears to be a mismatch between the type of question for which systematic reviews have been most successful and the type of question that is of greatest interest within energy policy and hence most likely to attract funding.

This suggests that energy researchers may only be able to use systematic review techniques for a subset of questions and may need to modify and extend those techniques when applied. In many cases, the conventional ‘narrative’ review is likely to be more appropriate. The greater use of systematic reviews will require a combination of increased awareness amongst researchers, appropriate training and changes in funding priorities. But while systematic reviews may have fundamentally changed the practice of medicine and health-care, there seems little prospect of a comparable impact on energy policy.

Acknowledgements

This research was funded by the UK Research Councils as part of the Technology and Policy Assessment (TPA) function of the UK Energy Research Centre. The author would like to thank Jim Skea (Research Director, UKERC), Rob Gross (Head of TPA, UKERC), Annette Boaz (Queen Mary College), Erik Millstone (SPRU Environment Programme, University of Sussex), Gordon Mackerron, Andy Stirling and John Dimitropoulos (all of the Sussex Energy Group, University of Sussex) for their helpful comments. The usual disclaimers apply.

References

- Alfredsson, E.C., 2004. Green consumption-no solution for climate change. *Energy*. **29**, 513-524.
- Archer, M., R. Bhaskar, A. Collier, T. Lawson and A. Norrie (Eds.), 1999. *Critical Realism: Essential Readings*, Routledge, London and New York.
- Bache, I.W., 2003. *Critical realism and econometrics*. Working Paper. Research Department, Norges Bank, Oslo.
- Barker, T., J. Koehler and M. Villena, 2002. The costs of greenhouse gas abatement: a meta-analysis of post-SRES mitigation scenarios. *Environmental Economics and Policy Studies*. **5**, 135-166.
- Bennett, T., D. Farrington and K. Holloway, 2004. *The effectiveness of neighbourhood watch: protocol for the Campbell Collaboration in Crime and Justice*. Centre of Criminology, University of Glamorgan, Pontypridd.
- Berndt, E.R. and D.O. Aard, 1979. Engineering and econometric interpretations of energy-capital complementarity. *American Economic Review*. **June**, 259-268.
- Bhaskar, R., 1975. *A Realist Theory of Science*. Leeds Books, Leeds.
- Blundell, R. and R. Walker, 2000. *Working Families' Tax Credit: a review of the evidence, issues and prospects for further research*. Centre for Economic Evaluation, Institute for Fiscal Studies, London.
- Boaz, A. and D. Ashby, 2003. *Fit for purpose? Assessing Research quality for evidence based policy and practice*. ESI UK Centre for Evidence Policy and Practice, Working Paper 11. Queen Mary College, University of London.
- Brookes, L., 2000. Energy efficiency fallacies revisited. *Energy Policy*. **28**(6-7), 355-366.
- Bullock, H., J. Mountford and R. Stanley, 2001. *Better policymaking*. Cabinet Office Centre for Management and Policy Studies, London.
- Cabinet Office, 2005. *The Magenta Book: guidance notes for policy evaluation*. London.
- Cooper, H. and L.V. Hedges (Eds.), 1994. *The Handbook of Research Synthesis*, Russell Sage Foundation, New York.
- Croucher, K., D. Quilgars, A. Wallace, S. Baldwin and L. Mather, 2003. *Paying the mortgage? A systematic literature review of safety nets for home owners*. Department of Social Policy and Social Work, University of York, York.
- Davies, A., T.O. Huw, S.M. Nutley and P.C. Smith, 2000. *What works: Evidence based policy and practice in the public services*. Policy Press, Bristol.
- Davies, H. and S.M. Nutley, 2002. *Evidence-based policy and practice: moving from rhetoric to reality*. Discussion Paper 2. University of St Andrews, St Andrews.
- Davies, H.T.O., S.M. Nutley and P. Smith, 1999. *What works? The role of evidence in public sector policy and practice*. *Public Money and Management*. **19**(1), 3-5.
- Davies, H.T.O., S.M. Nutley and P.C. Smith, 2000. *What works? Evidence-based policy and practice in public services*. Policy Press, Bristol.
- Davies, P.T., 1999. What is evidence-based education? *British Journal of Educational Studies*. **47**(2), 108-121.
- Department for Health, 1991. *Research for health: the research and development strategy for the NHS*. Department for Health, The Stationary Office, London.
- Espey, M., 1998. Gasoline demand revisited: an international meta-analysis of elasticities. *Energy Economics*. **20**(3), 273-295.
- Evans, J. and P. Benefield, 2001. Systematic reviews of educational research: does the medical model fit? *British Educational Research Journal*. **27**(5), 527-541.
- Florax, R.J.G.M., H.L.F. de Groot and R.A. de Mooij, 2002. *Meta-analysis: a tool for upgrading inputs of macroeconomic policy models*. Working Paper. Department of Spatial Economics, Vrije Universiteit, Amsterdam.
- Greening, L.A., Greene and C. Difiglio, 2000. Energy efficiency and consumption — the rebound effect — a survey. *Energy Policy*. **28**(6-7), 389-401.
- Grepperud, S. and I. Rasmussen, 2004. A general equilibrium assessment of rebound effects. *Energy Economics*. **26**(2), 261-282.

Improving the evidence base for energy policy

- Gross, R., 2005. User needs assessment: an analysis of policymaker requirements for evidence based policy assessment and implications for the TPA function of the UKERC. Imperial College Centre for Energy Policy and Technology, London.
- Hammersley, M., 2001. On 'systematic' reviews of research literatures: a 'narrative' response to Evans and Benefield. *British Educational Research Journal*. **27**(5), 543-554.
- Hammersley, M., 2005. Is the evidence-based practice movement doing more good than harm? Reflections on Iain Chalmers' case for research-based policy making and practice. *Evidence and Policy*. **1**(1), 85-100.
- Hand, M., D. Southerton and E. Shove, 2003. Explaining daily showering: a discussion of policy and practice. Sustainable Technologies Programme, Working Paper 2003/4. Centre for Research on Innovation and Competition, University of Manchester.
- Hodgson, G.M., 2004. On the Problem of Formalism in Economics. *Post-autistic Economics Review*. **28**.
- Hsieh, C. and M. Pugh, 1993. Poverty, income inequality and violent crime: a meta-analysis of recent aggregate data studies. *Criminal Justice Review*. **18**, 182-202.
- Kremers, H., P. Nijkamp and P. Rietveld, 1999. The scope of meta-analysis for transport policy impact analysis in environmental economics. Research Memorandum 1999-34. Faculty of Economics and Econometric is, Free University, Amsterdam.
- Lawson, T., 1997. *Economics and Reality*. Routledge., London.
- Levitt, R., 2003. GM crops and foods. Evidence, policy and practice in the UK: a case study. ESRC UK Centre for Evidence Based Policy and Practice Working Paper 20. Queen Mary College, University of London.
- Majone, G., 1997. From the positive to the regulatory state: causes and consequences of changes in the mode of governance. *Journal of Public Policy*. **17**(2), 139-167.
- Meyer, B.D., 1995. Natural and quasi experiments in economics. *Journal of Business and Economic Statistics*. **13**(2).
- Oakley, A., 2002. Social science and evidence based everything: the case of education. *Educational Revue*. **54**(3), 277-285.
- Oakley, A. and D. Fullerton, 1995. A systematic review of smoking prevention programmes for young people. EPPI Centre, London.
- Oliver, S., G. Peersman, A. Harden and A. Oakley, 1999. Discrepancies in findings from effectiveness reviews: the case of health promotion for older people in accident and injury prevention. *Health Education Journal*. **58**, 66-77.
- Peacock, M., 2000. Explaining theory choice: an assessment of the critical realist contribution to explanation. Cambridge Realist Workshop Conference, Cambridge.
- Petrosino, A., C. Turpin-Petrosino and J.O. Finckenaer, 2000. Well-meaning programmes can have harmful effects! Lessons from experiments of programmes such as Scared Straight. *Crime and Delinquency*. **46**(3), 354-379.
- Petticrew, M., 2001. Systematic reviews from astronomy to zoology: myths and misconceptions. *British Medical Journal*. **322**, 98-101.
- Petticrew, M. and H. Roberts, 2005. *Systematic Reviews in the Social Sciences: a practical guide*. Blackwell Publishing, Oxford.
- Pratschke, J., 2003. Realistic models? Critical realism and statistical models in the social sciences. *Philosophica*. **71**, 13-38.
- Roberts, K., M. Dixon-Woods, R. Fitzpatrick, K. Abrams and D. R. Jones, 2002. Factors affecting uptake of childhood immunisation: an example of Bayesian synthesis of qualitative and quantitative evidence. *The Lancet*. **360**, 1596-1599.
- Saunders, H.D., 2000. A view from the macro side: rebound, backfire, and Khazzoom-Brookes. *Energy Policy*. **28**(6-7), 439-449.
- Sayer, A., 1992. *Method in Social Science: a Realist Approach*. Routledge, London and New York.
- Schipper, L. and M Grubb, 2000. On the rebound? Feedback between energy intensities and energy uses in IEA countries. *Energy Policy*. **28**(6-7), 367-388.
- Small, K.A. and K. Van Dender, 2005. Study to Evaluate the Effect of Reduced Greenhouse Gas Emissions on Vehicle Miles Traveled. Final Report, ARB Contract Number 02-336. Department Economics, University of Irvine, Irvine, California.
- Solesbury, W., 2001. Evidence based policy: whence it came and where it's going. ESRC UK Centre for Evidence Policy and Practice: Working Paper 1. Queen Mary College, University of London, London.

Improving the evidence base for energy policy

- Sorrell, S., 2005. The role of evidence in energy policy: What can be learnt from Evidence Based Policy and Practice? Report for the Technology and Policy Assessment function of the UK Energy Research Centre. SPRU (Science and Technology Policy Research), Brighton.
- Sorrell, S. and J. Dimitropoulos, 2005. An assessment of evidence for rebound effect from improvements in energy efficiency: Scoping Note. Report for the Technology Policy Assessment function of the UK Energy Research Centre. SPRU (Science and Technology Policy Research), University of Sussex, Brighton.
- Spencer, L., J. Ritchie, J. Lewis and L. Dillon, 2003. Quality in quality evaluation: a framework for assessing research evidence. Government Chief Social Researcher's Office, The Cabinet Office, London.
- Sweeney, L. and C. Haney, 1992. The influence of race on sentencing: a meta-analytic review of experimental studies. *Behavioural Science Law*. **10**, 179-195.
- Wallace, A., K. Croucher, D. Quilgars and S. Baldwin, 2004. Meeting the challenge: developing systematic reviewing in social policy. *Policy and Politics*. **32**(4), 455-70.
- Willet, K.D. and S. Naghshpour, 1987. Residential demand for energy commodities : A household production function approach. *Energy Economics*. **9**(4), 251-256.