The use of evidence reviews to inform demand-side policy

Dr. Peter Warren
Department for Business, Energy and Industrial Strategy
UK Government
3 Whitehall Place
London, SW1A 2AW
UK

Keywords

policy, demand side management (DSM), evidence reviews, evidence quality

Abstract

The UK's Department for Business, Energy and Industrial Strategy (BEIS) undertakes and commissions programmes of research on the demand-side, managed by government scientists and analysts to inform and advise policy officials on the development of demand-side policies, such as energy efficiency, demand-side response and smart technologies. Evidence reviews, commonly confused with literature reviews, are a crucial method not only for collating and synthesising the evidence base, but also for determining what the quality of previous evidence is and extracting the most amount of value from previous studies using systematic techniques.

The paper focuses on the use of evidence reviews to inform the development of current and future demand-side management policies in the UK. An innovative framework and methodological approach for conducting evidence reviews for policy analysis and development is discussed and four recent government-commissioned research reports that have applied the approach are summarised as case studies: two focus on the international and UK evidence base for heating controls, one investigates the evidence base for occupancy patterns, and the fourth examines the evidence base for demand-side response. The evidence reviews have been used to inform the development of policies for domestic heat and smart energy. The paper concludes with recommendations on the use of evidence reviews to inform energy policy, and the application of UK experiences to other countries.

Introduction

Evidence is crucial for the development of government policy to ensure that policies maximise positive societal impacts, minimise any negative impacts, are defendable, and to ensure that new policies learn from past experiences, both domestically and internationally. Evidence reviews are a useful method for comprehensively collating and synthesising the evidence base on a particular topic. The method is commonly confused with literature reviews: the former is a method for collecting data for analysis, whereas the latter is not a method, but instead aims to critique the current literature and to identify research gaps. Evidence reviews are not widely used in the energy field, but Sorrell (2007) and Warren (2014a) have called for its greater use by adapting methods from other disciplines that use the method, such as education, crime and justice, and social welfare. Nevertheless, in the sub-field of energy policy analysis, the method is underutilised.

This paper argues that evidence reviews are a crucial method not only for understanding what has been done before, but also for determining what the quality of previous evidence is and extracting the most amount of value from previous studies using systematic techniques. They are a comprehensive and resource-efficient approach for governments to understand the evidence base and to utilise the results in policy development. This paper aims to answer the following research question: "How can evidence reviews be used to inform the development of demand-side policies?" by outlining a new framework to conceptualise the different types of evidence review and when they are most appropriate to use, and an innovative methodological technique to assess evidence quality, which has been applied to the areas of heating controls, occupancy patterns and demand-side response.

Section two provides background to the different forms of evidence, gives an overview of the different types of evidence review and briefly summarises how evidence reviews are conducted. Section three discusses the importance of evidence quality and provides criteria and a new practical scale for determining the quality of evidence. Section four outlines the new framework for understanding the different types and application of evidence reviews from a government perspective. Section five provides brief summaries of four case studies of how evidence reviews have been applied to inform the development of demand-side management (DSM) policies in the UK's Department for Business, Energy and Industrial Strategy (BEIS) (responsible for UK energy and climate policy), in particular policies for domestic heat and smart energy. Section six provides the conclusions of the paper.

Evidence reviews

Evidence refers to: "the available body of facts or information indicating whether a belief or proposition is true or valid" (Oxford Dictionary, 2017). There are various forms of evidence in the energy field across different disciplines, the most common being quantitative modelling (e.g. energy optimisation models, energy accounting models and energy simulation models), qualitative interviews (e.g. structured/semi-structured/unstructured interviews and focus groups) and surveys (e.g. quantitative/qualitative/mixed methods questionnaires). Less common are trials (e.g. randomised control trials and targeted (non-randomised) trials), ethnography (e.g. participant observation and shadowing), quantitative interviews (e.g. multi-criteria decision-making analysis and Q methodology) and evidence reviews (e.g. meta-analyses, realist syntheses and thematic summaries). There are various reasons for this pattern, though many revolve around methodological traditions in particular disciplines, a lack of expertise to apply methods from non-energy fields and resource costs (particularly for trials).

Although the appropriateness of the method is driven by the research questions, governments often require data and results that are nationally representative or statistically representative of particular targeted groups, such as those in fuel poverty (consumers that are "members of households that are living on a lower income in a home that cannot be kept warm at reasonable cost", as defined in the UK's *Warm Homes and Energy Conservation Act*, 2000) or a particular sector. Where the costs of achieving such samples are large, evidence reviews can provide a resource-efficient method for providing comprehensive data and analysis on a particular topic, as they involve collating and synthesising all of the work that has been done on a particular intervention, trial or programme to better understand what works and what does not (Petticrew and Roberts, 2006).

Evidence reviews, particularly systematic reviews (discussed in section four), are well established in the medical sciences, especially through the Cochrane Collaboration (providing a database of >5,000 systematic reviews), as well as in other disciplines, such as education and social policy, especially through the Campbell Collaboration, which was established in 2000 as the non-medical equivalent of the Cochrane Collaboration. In the energy field, the method is beginning to develop through initiatives, such as the UK's Collaboration for Environmental Evidence, the UK Energy Research Centre's Technology and Policy Assessment group, and the Research Councils UK's Centre for Energy Epidemiology (based in University College London). Examples include the impacts of energy systems on marine ecosystems (Papathanasopoulou *et al.*, 2016), the barriers to energy services in the poorest countries (Watson *et al.*, 2012) and bioenergy research (e.g. Muench and Guenther, 2013; Gurwick *et al.*, 2013; Rehfuess *et al.*, 2014). However, the use of evidence reviews in the energy policy sub-field remains limited.

In policy terms, evidence reviews can be defined as a method to comprehensively and systematically gather, collate and synthesise data to better understand the impacts of a policy or programme. There are various different methodological techniques that fall under this method, which have been comprehensively summarised in Dixon-Woods *et al.* (2005) and Snilstveit *et al.* (2012) from a methodological perspective, and in Sorrell (2007) and Warren (2014a) in relation to energy. Their

Table 1. The different types of evidence review (source: Warren, 2015).

Category	Systematic Review Type
Interpretive	Narrative Summaries
	Thematic Summaries
	Grounded Theory
	Meta-Ethnography
Integrative	Content Analysis
	Case Survey
	Qualitative Comparative Analysis Method
	Meta-Analysis
Mixed Methods	Meta-Study
	Realist Synthesis
	Miles and Huberman's Cross-Case Techniques
	Framework Synthesis
	Thematic Synthesis

reviews and explanations of different review techniques are not repeated here, but a review of these reviews by Warren (2015) is summarised in Table 1, which shows three over-arching methodological groups: integrative (primarily quantitative), mixed methods and interpretive (primarily qualitative).

In practice, evidence reviews follow the eight main stages below (adapted from Harden and Thomas, 2005):

1. Review questions and boundaries
2. Selection of evidence review type
3. Comprehensive search
4. Inclusion and exclusion criteria
5. Quality assessment scale
6. Data and information extraction
7. Synthesis of findings
8. Dissemination of findings

In summary, after the research questions, the review boundaries and the type of evidence review have been determined, a robust and replicable search strategy should be developed. This outlines the specific search terms and databases that will be used to search for evidence. The search terms should use Boolean logic to enhance their effectiveness, such as AND, OR, AND/OR, * and parentheses. A list of inclusion and exclusion criteria is necessary to reduce the scope of the review to what is feasible within the time and budget. Examples include: the languages, the geographical scope, the time horizon, the databases, the use of snowballing and hand searching. The assessment of evidence quality is the most important stage in the evidence review process and this is discussed in section three.

Once the final sample is obtained (following the quality assessment), a standardised process for extracting and recording data is required. A spreadsheet (or equivalent) is commonly used to do this. It is good practice to store the details of relevant evidence that did not pass the quality assessment stage, so that the breadth and focus of the evidence base as a whole can be examined. However, low quality evidence should not be included in the final sample or the main analysis. Once all of the data have been extracted in all of the documents in the sample, the data are analysed. In the case of integrative methodological techniques, data are usually statistically aggregated. However, for mixed methods and interpretive methodological techniques, data are usually synthesised. The latter is much more common in evidence reviews undertaken in the energy field due to the wide variety of methods used to collect evidence, which reduces the appropriateness of statistical aggregation (discussed in section three).

A crucial part of developing a robust review protocol (a document outlining stages 1–8 above) for evidence reviews is undertaking a pilot study to test the replicability and transparency of the methodological process. A minimum of three reviewers should independently use a small number of search terms in a few databases to ensure that the same documents are being retrieved and deemed relevant after the inclusion and exclusion criteria are applied. Following this, the reviewers should independently achieve the same quality assessment score for each document and produce the same final sample of documents. In practice, scores within ±1 of each other are acceptable, as the crucial aspect is that the same documents pass or fail the quality assessment stage and do not have scores that vary substantially within the 'pass' range. For example, if one reviewer scores a document maximum marks and another review just passes the same document on the threshold of pass or fail, the quality assessment process adopted needs to be reviewed to ensure its robustness and replicability. Section three discusses evidence quality assessment in more detail. The pilot study should then ensure that the same data are independently extracted for each document and stored using a standardised spreadsheet (or equivalent) template. The extracted data should be synthesised by each reviewer using the same methodological process in order to achieve the same results. Thus, the review protocol needs to be sufficiently detailed to allow those taking part in the pilot study to independently achieve the same results.

Assessing evidence quality

High quality evidence is crucial for underlying robust, defendable government policies. In the energy field, this is particularly challenging due to the diverse range of methods used, which vary in how they are conducted and what disciplinary perspective they are approached from. Furthermore, evidence from academia and industry also varies in how skilled experts are in understanding and communicating the policy implications of their work. In order to obtain a holistic understanding of the quality of evidence, it is important that a range of criteria are used to assess evidence quality. For example, peer review is an important indicator of evidence quality, but by itself, it does not give a comprehensive enough picture of quality and thus there is a need to use a broad range of criteria.

This section presents results from a review of the quality assessment criteria adopted in various disciplines that use evidence reviews. In particular, the practices of the medical sciences and social policy are examined, as these are the leading fields in applying two different types of evidence review (integrative in the former and interpretive in the latter). In addition, emerging (more mixed methods) practices in the energy field are included. Table 2 summarises the results.

Stage five of conducting evidence reviews involves assessing the quality of the retrieved evidence after studies have been deemed relevant in stage four. This filters the poor quality studies from the high quality studies in the production of the final sample. The quality assessment criteria shown in Table 2 are the more common metrics used in the disciplines that have a more established use of evidence reviews than in the energy policy sub-field. To assess all studies against all of the criteria shown in Table 2 would be resource-intensive and would unnecessarily penalise studies that excluded 'surface' reporting features, such as statements of conflicts of interest, copyright or regulatory compliance.

In practice, commonly used scales in other disciplines adopt 5–7 quality assessment criteria and have a threshold at which studies are deemed to be high quality. A threshold of two-thirds of the available points is frequently applied (i.e. for a scale with six quality criteria, a study would need to score four or more points to be included in the final sample). The review of quality assessment scales included: the Jadad Scale (1998 – medical sci-

ences), the guidelines of the Equator Network (Enhancing the Quality and Transparency of Health Research) and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) (medical sciences), the Pawson *et al.* TAPUPAS (Transparency, Accuracy, Purposivity, Utility, Propriety, Accessibility, Specificity) Framework (2003 – social care) and the Warren Scale (2014a; 2015 – energy policy).

The review led to the development of a new scale for use in energy policy, which is shown in Figure 1. The justification for a new scale is based on two central limitations of previous practices. Firstly, energy policy is a field that involves evidence from a diverse range of actors that have used different methods applied to different (but related) areas of focus, from different perspectives in various timescales in a range of geographical contexts. As such, other scales are less effective at dealing with this diversity of evidence. For example, the Jadad Scale (1998) is a useful and robust technique for evaluating randomised control trials (which are rare in the energy field) but it has few other applications. Furthermore, some critics, such as Berger (2006) and Clark *et al.* (1999), argue that the scale places too

Table 2. Criteria for assessing the quality of evidence.

Category	Group	Criteria
Reporting quality	Reporting errors	Type 1 error(s) – effect/relationship stated to exist when it does not
		Type 2 error(s) – effect/relationship stated not to exist when it does
		Type 3 error(s) – right answer(s) given to the wrong question(s)
		Transparency of data and datasets
		Totals given for percentages
		Study answers the research question(s)
		Conclusions match the data
		Specific to policy evaluation: Details on implementation process
		Specific to policy evaluation: Details on evaluation process
	Study outline	Clear and justified rationale and study aim(s)
		Clear and justified research question(s)
	Peer review	Peer reviewed
		Number of peer reviewers
		Expertise of peer reviewer(s)
		External independent peer reviewer(s)
		Internal independent peer reviewer(s)
		Internal or external non-independent peer reviewer(s)
	Resource contributions	Acknowledgement of funding
		Acknowledgement of other resource contributions
		Statement(s) of conflicts of interest
		Statement(s) of copyright
		Statement(s) of regulatory compliance
Research quality	Expertise	Expertise of author(s)
		Track record of author(s)
		Track record and expertise of author(s)'s institution
		Track record and expertise of publishing institution
	Design	Study is legal
		Study is ethical
		Accessible to targeted audience(s)
		Meets the quality standards already used for the type of knowledge
	Methodological quality	Method(s) appropriate for answering the research question(s)
		Method(s) undertaken correctly
		Specific to certain methods: Blinded study
		Specific to certain methods: Appropriate method for blinding
		Specific to certain methods: Randomised study
		Specific to certain methods: Appropriate method for randomisation
		Specific to certain methods: Withdrawal rates stated

much emphasis on blinding, can show low consistency between researchers and is over-simplistic. The Cochrane Collaboration has also argued that the Jadad Scale puts too much emphasis on research reporting rather than research conduct (2011). At the other end of the spectrum, the Pawson *et al*. TAPUPAS Framework (2003) is broad enough to usefully cover a wide variety of methodological types, but it introduces a large degree of judgement on the part of the assessor, thus potentially reducing its reliability in consistently achieving independent replication by different assessors. Furthermore, it excludes what this paper argues should be key quality assessment criteria in any scale: peer review and matching data to conclusions.

The strength of Figure 1 is that it covers a broad range of key quality criteria across both reporting quality and research quality, and it can be applied to a wide variety of evidence in terms of research methods, context, time horizon and geographical coverage. The scale allows all relevant evidence to be assessed equally and studies that score two-thirds of the total number of points (six or more points out of a total of nine points) are deemed high quality. The review highlighted that an important limitation of currently available scales in other disciplines is that they focus primarily on research quality rather than reporting quality. This paper argues that the two are equally important as reporting quality ensures that studies are transparent, reliable and can be replicated.

Although the scale was developed to inform the development of DSM policies (activities on the demand-side of energy meters – see Warren, 2014b for a full definition), the adoption of broad criteria enable it to be applied in any discipline or project where the focus is to synthesise a diverse evidence base. However, a 'diverse' evidence base should not be confused with a 'weak' evidence base. An evidence review is inappropriate if the evidence base in a particular area is weak. As such, it is common for a literature review to be conducted prior to an evidence review to scope out the strength of the evidence base, in order to determine the suitability of an evidence review for answering particular research questions.

This paper argues that the weaknesses of any one particular quality criterion are offset by the use of a range of different criteria in scales. For example, in Figure 1, the track record of the author(s) can be an indicator of the consistency of evidence quality. However, by itself, the criterion is limiting, as it excludes new authors in the field that might have produced high quality work. Thus, by utilising a range of criteria across both research and reporting quality the reliability of the scale is increased, as studies are not penalised if they do not obtain the point for every criterion available.

In some scales, weightings are used where specific criteria are given more points if they are deemed by the reviewers to be more important in the context of a specific project. In Figure 1, the scale gives equal weighting to all of the criteria, but groups them into similar areas of assessment (e.g. for rationale and research questions, the rationale for the study might be clear and justified, but the research questions might be poorly constructed in meeting the research aims). Some scales, such as the Jadad Scale (1998) give additional points for certain criteria. For example, for the criterion: "Was the study described as randomised?", an additional point is given if the method of randomisation was appropriate and a point is deducted if the method was inappropriate. The more simplistic approach of the scale shown in Figure 1 ensured that it has wide applications across energy policy areas where the strength of the evidence base varies. For example, in section five, the four case studies utilised the scale, but the strength of the evidence base varied: for occupancy patterns, the evidence base is weak at the generalisable level, for heating controls, it is varied (though generally weak), and for demand-side response, it is much stronger.

The use of scales to assess evidence quality is the most important part of evidence reviews. However, quality assessment scales have much wider applications and can be used to assess the quality of submitted evidence (evidence provided directly by stakeholders or experts), or to inform the design of primary evidence collection and reporting to ensure that new evidence is high quality.

Evidence reviews in government

The Government Social Research profession of the UK Civil Service has acknowledged the use of evidence reviews to inform government policy, but the current guidance is arguably out-of-date and is currently archived, thus limiting the development of evidence review expertise (see 'Rapid Evidence Assessment Toolkit': http://webarchive.nationalarchives.gov.uk/20140305122816/http:/www.civilservice.gov.uk/networks/gsr/resources-and-guidance/rapid-evidence-assessment). There is also a strong focus on just one type of evidence review, the Rapid Evidence Assessment (see below). However, evidence reviews range from reviews of reviews to systematic reviews with the type of review determined by the resources required to undertake it, such as time, the number of reviewers and the budget. This paper presents a new framework for understanding the different types of evidence reviews from the perspective of governments that commission them, which is shown in Figure 2. Figure 2 also includes

- 2 points: Are the rationale (1 point) and research questions (1 point) clear and justified?
- 2 points: Does the document acknowledge resource contributions (1 point) and possible conflicts of interest (1 point)?
- 2 points: Has the document been peer reviewed or independently verified by one (1 point) or more (2 points) reputable experts?
- 1 point: Are the methods used suitable for the aims of the study?
- 1 point: Do the conclusions match the data presented?
- 1 point: Does the author / publishing organisation have a track record in the area?

Figure 1. BEIS Quality Assessment Scale.

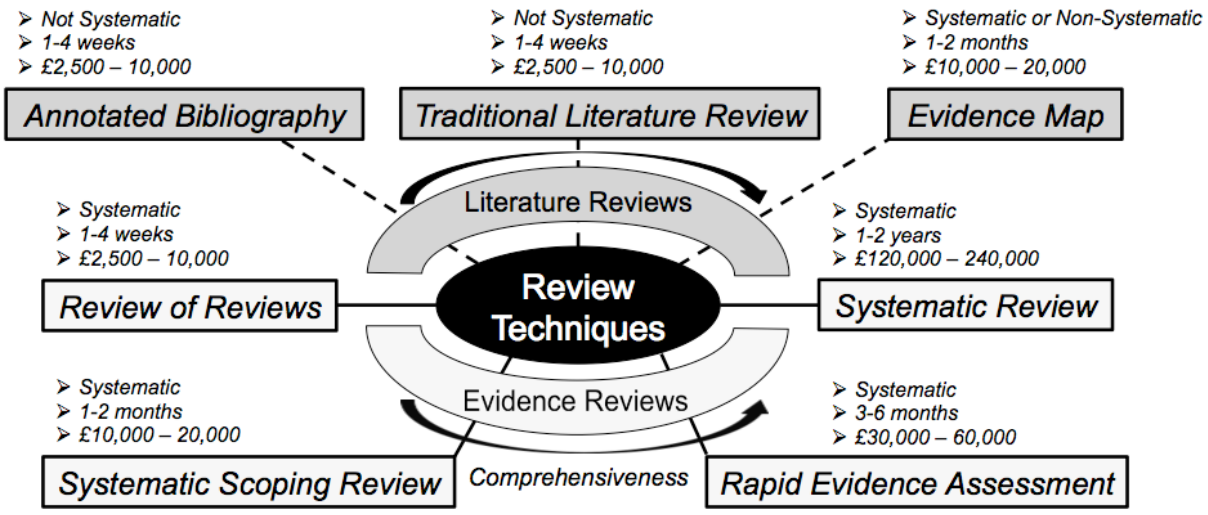

Figure 2. Evidence Reviews Framework.

other review techniques that do not come under the umbrella term of 'evidence reviews', such as annotated bibliographies, traditional literature reviews and evidence maps, which are not methods for collecting data for analysis and usually do not use systematic techniques. The paper does not discuss these forms of review techniques but instead focuses on the use of evidence reviews to inform demand-side policy. Evidence reviews are appropriate for any policy area where the evidence base is sufficient enough.

The evidence reviews framework is developed from both a (literature) review of evidence review practices in academia (as discussed in section three) and from the practical application of new techniques, such as Figure 1, to government-commissioned projects conducted in 2016 (which are discussed in section five). The costs are based on UK experiences and may not be reflective of the costs in other countries due to differences in context. As Figure 2 shows, there are four main types of evidence review: reviews of reviews, systematic scoping reviews, rapid evidence assessments (REAs) and systematic reviews. In Figure 2, as the reader moves from left to right of the diagram, the comprehensiveness (and resource requirements) of the review increases for both evidence reviews and literature reviews. Indicative costs (in UK pounds) and timescales are provided based on the experiences of BEIS.

Reviews of reviews are the least comprehensive type of evidence review, but they are the most time-efficient. They use systematic techniques and reduce the scope of the review by just focusing on review studies (usually literature reviews), which have already brought together the primary evidence on a particular topic. Data are then extracted from the review studies directly, and where required, the original studies may be retrieved in order to obtain the primary data. A drawback of this type of evidence review is that it relies on the existence of other reviews, thus potentially excluding primary evidence that has not yet been included in reviews. As such, the conclusions that can be drawn on the nature of the evidence base as a whole is limited. Nevertheless, reviews of reviews can produce results more quickly, more transparently and more reliably than non-systematic literature review techniques, and can be used to answer both broad and narrow research questions.

Systematic scoping reviews are resource-efficient evidence reviews that can provide a comprehensive synthesis of the evidence base in a short period of time at low cost. They are usually 1–2 months in duration and are systematic. However, in certain circumstances, non-systematic scoping reviews may be conducted where the review is based on submitted evidence rather than through the active searching of databases and evidence sources using search strategies and inclusion criteria. Nevertheless, submitted evidence should still be subjected to a quality assessment scale, as discussed in section three. This paper advocates the use of systematic scoping reviews as the most comprehensive and efficient type of evidence review for governments to use based on time and budget constraints (demonstrated in section five). However, systematic scoping reviews are suitable for answering a few, narrow research questions rather than broader questions, as this is a necessary requirement for reducing the scope of the review but whilst maintaining a comprehensive synthesis of the evidence base.

Rapid Evidence Assessments (REAs) are the most commonly commissioned type of evidence review in the energy field in the UK (whether commissioned by government or by other funding bodies). Despite being more resource-intensive than systematic scoping reviews (in terms of time and budget), they comprehensively review the entire evidence base on a particular topic and can be used to answer a much broader array of research questions than the other types of evidence review (with the exception of systematic reviews). In setting up an REA, the process for planning a systematic review is followed and then the review boundaries are reduced by limiting the number of search terms and databases that are examined. Thus, unlike other types of evidence review, a literature review of the main databases and data sources that produce evidence on a given topic is first required. In contrast, for systematic scoping reviews and reviews of reviews, the choice of databases and data sources is usually targeted based on the knowledge of the reviewers and the commissioning body. An example of an REA conducted to inform the development of DSM policies in the UK can be found here: https://www.gov.uk/government/publications/how-heating-controls-affect-domestic-energy-demand-a-rapid-evidence-assessment. The REA examined how

heating controls affect domestic energy demand (Department of Energy and Climate Change (DECC), 2014).

In the medical sciences, systematic reviews form the top of the *Hierarchy of Evidence* pyramid (developed by the University of Illinois and discussed in Warren, 2014a), followed by randomised control trials, cohort studies, case control studies, case series/case reports and then editorials/expert opinions. They are resource-intensive but are fully comprehensive and usually include a large number of search terms, a large number of databases and data sources, a wide-ranging list of inclusion and exclusion criteria (such as including snowballing, hand-searching and non-English documents), and are conducted by a larger team of reviewers. As Figure 2 shows, systematic reviews take a minimum of one year to complete, which is usually beyond the timescales of most governments, as they require evidence to inform the development of policies at the correct stage in the policy process. In the UK, this occurs at the call-for-evidence stage of policy development, though it is not uncommon for evidence to be submitted before this period. However, after this period, the impact of submitted evidence is more limited, as policies would have entered the policy implementation stage of the policy process. Systematic reviews are usually beyond the resources of governments and still remain a method that has had limited application in the energy policy field (however, Warren (2015) is an example of the development of a methodological approach to apply systematic review methodology to the energy policy field).

Evidence review expertise in the energy field is limited; however, the use of the method is growing in BEIS to inform the development of DSM policies, such as domestic heat and smart energy. Certain techniques that evolved through the methodological advancement of evidence reviews in academia, such as realist evaluation (which developed from realist synthesis systematic review principles – see Pawson and Tilley, 1997; Pawson, 2002a; 2002b; 2006; Warren, 2015; Dixon-Woods *et al.*, 2006; and the RAMESES (Realist And Meta-narrative Evidence Syntheses: Evolving Standards) project), are also beginning to be applied in energy and climate policy evaluation. Furthermore, there have been methodological developments within BEIS to look at the integration of evidence reviews, particularly systematic scoping reviews and REAs, with other methods, such as interviews, case studies and non-systematic evidence searching. These developments allow new primary data to be combined with a comprehensive review of existing data to ensure that policies are developed in an up-to-date context, and are not based on out-dated evidence.

Despite the strength of evidence reviews as a robust and comprehensive method, it has some drawbacks. Firstly, due to its limited use in the energy field to date, it is methodologically underdeveloped when applied to energy issues, such as demand-side policy. For example, evidence reviews that can combine studies that have been undertaken in different contexts using different methods are still being developed. This is in contrast to other fields, such as the medical sciences, where randomised control trials (RCTs) can be statistically aggregated as they have been subjected to the same process and contexts. However, RCTs are rare in the energy field. Examples of new techniques that can be applied to energy are emerging, such as the four demand-side policy case studies discussed in section five and Warren (2015). Secondly, evidence reviews are usually time-consuming and resource-intensive, and require specific evidence review expertise that is currently lacking in the energy field. Despite this, the development of systematic scoping reviews and REAs contribute to alleviating some of the concerns regarding resources.

Demand-side policy case studies

This section gives brief summaries of four government-commissioned evidence reviews that were commissioned and completed in 2016–2017 to inform various DSM policies, such as domestic heat, energy efficiency and demand-side response. Two of the evidence reviews were systematic scoping reviews of the UK evidence base and the international evidence base on domestic heating controls, building on the REA published in 2014 (DECC, 2014 – highlighted in section four), to inform domestic heat policies. The latter employed non-systematic evidence gathering and document analysis alongside the evidence review, though did not integrate the two methods during analysis procedures ('mixed-methods' analysis, as defined in Saunders *et al.*, 2009). A third evidence review was a systematic scoping review of the evidence base on domestic occupancy patterns in order to inform not only the development of policies for domestic heat and energy efficiency, but also to lay the foundations for a large-scale primary evidence-gathering project. The fourth evidence review was an REA of the UK and international evidence base on demand-side response, but which also undertook in-depth country case studies and integrated the two methods at the synthesis stage of review protocol (stage seven, as per the list of stages given in section two) ('mixed-model' analysis, as defined in Saunders *et al.*, 2009).

The systematic scoping review of the UK evidence base on heating controls was commissioned and published in 2016. The review focussed on determining the strength of the evidence base for a range of heating control types: central timers, room thermostats, programmable thermostats, thermostatic radiator valves (TRVs), time proportional and integral controls (TPIs), weather compensators, automation (including self-learning), zonal control, and remote control. The results and details of the review are provided in BEIS (2016a) and are not discussed in this paper. The varied strength of the evidence for heating controls, broken down by heating control type, was useful in informing the development of a domestic heat policy within the correct timescales for the policy (during the pre-call-for-evidence stage of policy development). The design of a comprehensive review that could be completed within one month to address an important policy need contributed to increasing the policy impact of the review.

The success of the review led to the commissioning of a more expanded two-month systematic scoping review that examined the international evidence base on heating controls. The second review used the same review protocol as the first review but with an increased number of search terms, data sources and inclusion criteria (such as including non-English documents in Portuguese, Spanish and German based on the linguistic capabilities of the review team), and an expanded list of heating control types: weather compensation, TPI controls, zonal control, programmable TRVs, manual TRVs, learning algorithms, automation, optimisation, modulating room (or load compensating) thermostats, communication protocols, remote con-

trol (such as via an App), occupancy sensors, programmable thermostats, on/off switches, boiler thermostats, central timers, room thermostats, geolocation, geofencing, and hot water controls. The review was commissioned in 2016 and published in 2017, and the results and details of the review are provided in BEIS (2017a), as it is beyond the focus of this paper to discuss the results.

The international review was split into two parts – the first part was a systematic scoping review and the second part was a non-systematic 'call-for-evidence' to stakeholders in different countries, notably Germany, Chile, Australia, Portugal and the USA (the justification for the inclusion criteria is provided in BEIS, 2017a). The focus of the second part was to review the policy, regulatory and legislative experiences of other countries with regards to heating controls. As such, integrating high quality submitted documents into the systematic scoping review in part one was inappropriate due to the different focus of each part. However, where the focus is the same for a given project, integrating submitted evidence into an evidence review is an acceptable methodological approach so long as it it is included in the list of inclusion criteria, and the documents are subjected to the same screening process for relevance and quality.

The systematic scoping review of occupancy patterns was commissioned and published in 2016, and the results are provided in BEIS (2016b). The aim of the review was to examine the evidence base for domestic occupancy patterns, including reviewing the more common methods for determining occupancy patterns, examining the relationship between occupancy patterns and heating patterns, and identifying key evidence gaps. The review was not only used to inform domestic heat policies, but to contribute to the design of a large government-commissioned data collection project, as robust, representative real-world data are crucial for policy design and implementation. The review highlighted that the evidence base for occupancy patterns is currently not generalisable to the UK population and few studies sought to determine different categories of occupancy patterns. Heating patterns appeared to be dependent on occupancy patterns in the UK context, which is important for understanding the potential impact of smart technologies, such as occupancy-based smart heating controls. The methodological process for this review draws parallels to that adopted for the systematic scoping review of the UK evidence on heating controls, as just one method was adopted.

The REA of the evidence base on demand-side response was commissioned in 2016 and published in 2017, and the results are provided in BEIS (2017b). The aim of the REA was to provide a robust evidence base for the development of demand-side response policies for small energy users up to 2025. The five-month review covered the domestic sector and SMEs (small-to-medium-sized enterprises), and was global in scope. The REA was integrated with five in-depth country case studies on Finland, Germany, ERCOT (Texas), Norway and PJM (Illinois). The case studies involved five semi-structured interviews with key stakeholders in each of the countries (such as regulators, system operators, utilities, aggregators and consumer organisations) and document analysis (such as documents submitted by the interviewees and from non-systematic evidence gathering). The purpose of the case studies was to understand the reasons behind the successes and failures of demand-side response markets that are more mature than the market in the UK. This level of depth complemented the findings of the REA, as evidence reviews of this nature tend to focus across contexts, methods and evidence sources to look at the evidence base as a whole, rather than focussing on specific countries or contexts.

A key part of the analysis for both the REA and the case studies was to identify the applicability of international experiences to the UK context. The selection of the case study countries was determined following a contextual analysis of countries that have a similar context to the UK (such as market structure), which is described in BEIS (2017b). Unlike the systematic scoping review of the international evidence base on heating controls, which also utilised a second method for data collection, the REA integrated the two methods. The integration took place at the synthesis stage of the review protocol (stage seven), where the results from both methods were analysed separately first and then analysed together (a mixed-model approach). The results contributed to the design of smart energy policies in the UK, particularly feeding in at the post-call-for-evidence stage of policy development in the design of policy proposals for demand-side response. The overall key findings from the four evidence reviews are summarised in Table 3.

Table 3. Overall key findings from the four demand-side policy case studies.

Evidence Review	Key Findings
Scoping Review of Heating Controls	Limited evidence base and limited robust evidence. Energy savings, cost-effectiveness and usability of heating controls are inconclusive. Some robust evidence for energy savings from room thermostats, time proportional and integral controls, thermostatic radiator valves, zonal control and automation (including learning).
Heating Controls: International Evidence Base and Policy Experiences	Limited evidence base and limited robust evidence. Energy savings, cost-effectiveness and usability of heating controls are inconclusive. Few countries have mandatory regulations or voluntary policies to encourage standard or advanced controls.
Scoping Review of Occupancy Patterns	Limited evidence base to determine the energy savings from occupancy-based smart heating controls. The most common method for measuring occupancy is time-use surveys but the most efficient method is sensor networks. Heating patterns appear to be strongly dependent on occupancy patterns.
Demand-Side Response Rapid Evidence Assessment	*[Note – results are forthcoming and will be published by the time of the conference]*

Conclusion

Evidence reviews, commonly confused with literature reviews, are a crucial method not only for collating and synthesising the evidence base, but also for determining what the quality of previous evidence is and extracting the most amount of value from previous studies using systematic techniques. The paper focused on the recent experiences of the UK's Department for Business, Energy and Industrial Strategy (BEIS) in commissioning and using evidence reviews to inform the development of current and future demand-side management (DSM) policies in the UK, such as policies for domestic heat and smart energy.

The paper aimed to answer the following research question: "How can evidence reviews be used to inform the development of demand-side policies?" In order to do this, the paper proposed a new technique for assessing the quality of evidence in the energy policy field, provided a framework through which government, industry and academia can understand the practicalities of different types of evidence review, particularly from a government perspective, and briefly discussed four government-commissioned evidence reviews in 2016–2017 that have applied the technique and framework as case studies. Two of the case studies focus on the UK evidence base and the international evidence base for heating controls, a third investigates the evidence base for occupancy patterns, and the fourth examines the evidence base for demand-side response.

This paper advocates the use of systematic scoping reviews and Rapid Evidence Assessments (REA) as resource-efficient and comprehensive methods for informing the development of DSM policies within the timescales and resources of governments. They ensure that only high quality evidence is used, thus helping to ensure that policies maximise positive societal impacts, minimise any negative impacts, are defendable from an expert perspective, and that they learn from past experiences, both domestically and internationally. The practical quality assessment scale and evidence reviews framework can be readily applied to other contexts, countries and disciplines, and their use is encouraged in other governments, as well as in academia and industry. Despite this, there still remain important skills and knowledge gaps, such as the need to develop evidence reviews expertise in energy policy analysis, continuing to adapt the evidence review practices of other disciplines to energy policy, and for academic and industry researchers to better understand and communicate the policy implications of their work, which is tailored to the timescales of specific policies. Evidence reviews have an important role to play in informing the development of energy policies.

Bibliography

BEIS (UK Department for Business, Energy & Industrial Strategy): www.gov.uk/government/organisations/department-for-business-energy-and-industrial-strategy.

BEIS (2016a) "Scoping Review of Heating Controls Project", www.gov.uk/government/uploads/system/ uploads/attachment_data/file/573888/ Final_Report_-_Heating_Controls_Scoping_Review_Project.pdf.

BEIS (2017a) "Heating Controls: International Evidence Base and Policy Experiences", www.gov.uk/government/policies/household-energy.

BEIS (2016b) "Occupancy Patterns Scoping Review Project", www.gov.uk/government/uploads/system/ uploads/attachment_data/file/565281/Final_ Report_-_Occupancy_Patterns_Scoping_Review_Project.pdf.

BEIS (2017b) "Realising the Potential of Demand-Side Response to 2025", publication forthcoming.

Berger, V.W. (2006) "Is the Jadad score the proper evaluation of trials?", The *Journal of Rheumatology*, 33 (8), pp. 1710–1712.

Campbell Collaboration website: http://www.campbellcollaboration.org.

Clark, H.D., Wells, G.A., Huët, C., McAlister, F.A., Salmi, L.R., Fergusson, D. and Laupacis, A. (1999) "Assessing the quality of randomised trials: reliability of the Jadad scale", *Controlled Clinical Trials*, 20 (5), pp. 448–452.

Cochrane Collaboration website: http://www.cochrane.org/.

Collaboration for Environmental Evidence (CEE) website: http://www.environmentalevidence.org.

DECC (UK Department of Energy and Climate Change) (2014) "How heating controls affect domestic energy demand", www.gov.uk/government/publications/how-heating-controls-affect-domestic-energy-demand-a-rapid-evidence-assessment.

Dixon-Woods, M., Agarwal, S., Jones, D., Young, B. and Sutton, A. (2005) "Synthesising qualitative and quantitative evidence – a review of possible methods", *Journal of Health Services Research Policy*, 10 (1), pp. 45–53.

Equator Network (Enhancing the Quality and Transparency of Health Research) guidelines: http://www.equator-network.org/?post_type=eq_guidelines&eq_guidelines_study_design=systematic-reviews-and-meta-analyses&eq_guidelines_clinical_specialty=0&eq_guidelines_report_section=0&s=+.

Gurwick, N.P., Moore, L.A., Kelly, C. and Elias, P. (2013) "A systematic review of biochar research, with a focus on its stability in situ and its promise as a climate mitigation strategy", *PLoS ONE*, 8 (9).

Harden, A. and Thomas, J. (2005) "Methodological issues in combining diverse study types in Systematic Reviews", *International Journal of Social Research Methodology*, 8 (3), pp. 257–271.

Jadad, A. (1998) *Randomised controlled trials: a users guide*, London: BMJ Books, UK.

Muench, S. and Guenther, E. (2013) "A systematic review of bioenergy life cycle assessments", *Applied Energy*, 112, pp. 257–273.

Oxford Dictionary (2017): https://en.oxforddictionaries.com/definition/evidence.

Papathanasopoulou, E., Queirós, A.M., Beaumont, N., Hooper, T. and Nunes, J. (2016) "What are the local impacts of energy systems on marine ecosystem services", Collaboration for Environmental Evidence.

Pawson, R. (2002a) "Evidence-based policy – in search of a method", *Evaluation*, 8 (2), pp. 157–181.

Pawson, R. (2002b) "Evidence-based policy: the promise of Realist Synthesis", *Evaluation*, 8 (3), pp. 340–35.

Pawson, R. (2006) *Evidence-based Policy: A Realist Perspective*, London: SAGE, UK.

Pawson, R. and Tilley, N. (1997) *Realistic Evaluation*, London: SAGE, UK.

Pawson, R., Boaz, A., Grayson, L., Long, A. and Barnes, C. (2003) "SCIE Knowledge review 03: Types and quality of knowledge in social care", *Social Care Institute for Excellence (SCIE)*, November 2003.

Petticrew, M. and Roberts, H. (2006) *Systematic Reviews in the Social Sciences*, Blackwell Publishing, Oxford, UK.

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines: http://www.prisma-statement.org.

RAMESES project.

RCUK Centre for Energy Epidemiology: http://cee.ac.uk.

Rehfuess, E.A., Puzzolo, E., Stanistreet, D., Pope, D. and Bruce, N.G. (2014) "Enablers and barriers to large-scale uptake of improved solid fuel stoves: a systematic review", *Environmental Health Perspectives*, 122 (2), pp. 120–130.

Saunders, M., Lewis, P. and Thornhill, A. (2009) *Research Methods for Business Students*, 5th Edition, Harlow: Pearson Education Limited.

Snilstveit, B., Oliver, S. and Vojtkova, M. (2012) "Narrative approaches to systematic review and synthesis of evidence for international development policy and practice", *Journal of Development Effectiveness*, 4 (3), pp. 409–429.

Sorrell, S. (2007) "Improving the evidence base for energy policy: the role of systematic reviews", *Energy Policy*, 35, pp. 1858–1871.

UK Civil Service, Rapid Evidence Assessment Toolkit, http://webarchive.nationalarchives.gov.uk/20140305122816/http:/www.civilservice.gov.uk/networks/gsr/resources-and-guidance/rapid-evidence-assessment.

UK Energy Research Centre (UKERC)'s Technology and Policy Assessment group: http://www.ukerc.ac.uk/programmes/technology-and-policy-assessment.html.

University of Illinois, *Levels of evidence*, can be downloaded from: http://ebp.lib.uic.edu/nursing/node/12.

Warm Homes and Energy Conservation Act of 2000, London: HMSO, UK.

Warren, P. (2014a) "The use of systematic reviews to analyse demand-side management policy", *Energy Efficiency*, 7, pp. 417–427.

Warren, P. (2014b) "A review of demand-side management policy in the UK", *Renewable and Sustainable Energy Reviews*, 29, pp. 941–951.

Warren, P. (2015) *Demand-Side Management Policy: Mechanisms for Success and Failure*, PhD Thesis, University College London (UCL), UK.

Watson, J., Byrne, R., Morgan Jones, M., Tsang, F., Opazo, J., Fry, C. and Castle-Clarke, S. (2012) "What are the major barriers to increased use of modern energy services among the world's poorest people, and are interventions to overcome these effective?" Project Report, Collaboration for Environmental Evidence.