Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field

Nicholas Mays, Catherine Pope¹, Jennie Popay²

Health Services Research Unit, Department of Public Health and Policy, London School of Hygiene and Tropical Medicine, London; ¹School of Nursing & Midwifery, University of Southampton, Southampton; ²Institute for Health Research, University of Lancaster, Lancaster, UK.

Policy-makers and managers have always used a wide range of sources of evidence in making decisions about policy and the organization of services. However, they are under increasing pressure to adopt a more systematic approach to the utilization of the complex evidence base. Decision-makers must address complicated questions about the nature and significance of the problem to be addressed; the nature of proposed interventions; their differential impact; cost-effectiveness; acceptability and so on. This means that Cochrane-style reviews alone are not sufficient. Rather, they require access to syntheses of high-quality evidence that include research and non-research sources, and both qualitative and quantitative research findings.

There is no single, agreed framework for synthesizing such diverse forms of evidence and many of the approaches potentially applicable to such an endeavour were devised for either qualitative or quantitative synthesis and/or for analysing primary data. This paper describes the key stages in reviewing and synthesizing qualitative and quantitative evidence for decision-making and looks at various strategies that could offer a way forward. We identify four basic approaches: narrative (including traditional ‘literature reviews’ and more methodologically explicit approaches such as ‘thematic analysis’, ‘narrative synthesis’, ‘realist synthesis’ and ‘meta-narrative mapping’), qualitative (which convert all available evidence into qualitative form using techniques such as ‘meta-ethnography’ and ‘qualitative cross-case analysis’), quantitative (which convert all evidence into quantitative form using techniques such as ‘quantitative case survey’ or ‘content analysis’) and Bayesian meta-analysis and decision analysis (which can convert qualitative evidence such as preferences about different outcomes into quantitative form or ‘weights’ to use in quantitative synthesis).

The choice of approach will be contingent on the aim of the review and nature of the available evidence, and often more than one approach will be required.

Introduction

Reviews for policy and management

Beginning in medicine in the 1980s, pressure to strengthen the link between evidence, and policy and practice developments has now spread across the field of public policy. To date, the dominant question driving systematic reviews of evidence has been effectiveness, or ‘what works?’, with the result that methods for systematically reviewing ‘quantitative’ research about effectiveness, notably research using experimental or quasi-experimental designs, have developed rapidly, targeted predominantly at the needs of clinicians.¹ ³ However, alongside these developments (and impressive as they have been), there has also been growing unease about the utility of the current generation of systematic reviews as an aid to decision-making in specific settings, particularly for policy-makers, managers and patients.

There is now widespread recognition that a review of evidence aiming to support the complex and often messy decision-making that policy-makers and managers are involved in will of necessity have to address a wider range of questions at different points in the decision-making process, including: in what sense is this a ‘problem’; how and why has it come about; how significant is it compared with other problems on the policy or management agenda; how is it changing over time; what is likely to work to address the problem, for
whom and under what circumstances; what if any are the differential impacts of interventions across social groups; how cost-effective are different policy or management options likely to be in the current context; and how acceptable are different interventions to target groups, the wider population and providers?

In answering such questions, it is clear that policy-makers and managers will often need to draw on diverse sources of evidence – not only quantitative and qualitative research but also other evidence such as expert opinion and explicit value judgements (e.g. public preferences as to the weight to be given to different policy outcomes). For example, recent research in Alberta, Canada, shows that health authority decision-makers take a broad and pragmatic view of what constitutes relevant evidence for priority-setting, including non-randomized quantitative studies, qualitative research, expert opinion and other more subjective sources such as anecdotal reports. Of course, this does not absolve decision-makers from the responsibility of being critical about the quality of the evidence they use and systematic in how they gather it, but it suggests that it is at least as important that they are explicit about the value criteria used.

Qualitative methods are increasingly used alongside quantitative methods in primary research on the grounds that the two sets of data can be complementary. The same logic potentially applies in systematic reviews, but is rarely applied. Methods for incorporating qualitative research and other evidence in systematic reviews are relatively under-developed, and present a major methodological and practical developmental challenge.

Aims of the paper

This paper aims to provide for commissioners and researchers:

- a description of the main stages in a systematic review of evidence from research and non-research sources designed to inform decision-making by policy-makers and managers;
- an indication of the range of evidence that could potentially be incorporated into such reviews;
- pragmatic guidance on the main methodological issues that will need to be addressed, given the early stage of development of methods for such reviews and with a particular focus on the synthesis of qualitative and quantitative evidence;
- an introduction to some of the approaches available to synthesize these different forms of evidence; and
- an indication of the types of review questions particular approaches to synthesis are best able to address.

Some definitions

The terms 'review' and 'synthesis' are often used synonymously and confusingly. For this paper, review is used to refer to the whole process of bringing together a body of evidence which can be drawn from research and other sources, relevant to a particular decision in a policy or management context. The term synthesis is used to refer to the stage of a review when the evidence extracted from the individual sources is brought together in some way. This may entail simply juxtaposing findings from different sources or identifying and extracting common themes across sources or integrating data from several sources to produce new insights or theories.

Skills for policy and management relevant reviews

If the relevant evidence is diverse, it follows that researchers may not always be best placed to interpret non-research sources of information in the context of a specific decision. For example, understanding the likely strategic reactions of key stakeholders may be undertaken best by those responsible for stakeholder management in an organization. Equally, the research skills required to synthesize research evidence are different from those needed to establish a consensus judgement on an issue across a range of ‘experts’. This highlights the need for a broad range of skills and experience in the team when undertaking policy and management reviews.

Is it feasible to synthesize disparate evidence?

There is controversy about the legitimacy and feasibility of combining the findings of research studies that use different methods. This controversy is especially pronounced with regard to qualitative research, where some have suggested that attempts at aggregation destroy the integrity of individual studies. This critique is informed by a ‘relativist’ position which argues that research offers multiple ‘truths’ or realities such that each study represents a unique, personalized view that cannot be replicated or transferred. This is further complicated when a synthesis attempts to integrate research using different methods or studies informed by different theories of knowledge. Relativists would suggest that differences in theory and method are fundamental and militate against integration of research. Against this, the view characterized as ‘subtle realism’ suggests that, while there may well be multiple descriptions or explanations of phenomena, these ultimately relate to some underlying reality or truth. From this realist perspective, which we adopt, synthesis is accepted as promoting a greater understanding.

These controversies are writ large when we come to consider the synthesis of qualitative and quantitative evidence. There are those who contend that the divide between the qualitative and quantitative paradigms is far greater than the differences between different types of qualitative research, and that it is an irremovable barrier to synthesis. However, to some extent this
divide between qualitative and quantitative research is being bridged as many researchers now engage in multi-method research, utilizing both quantitative and qualitative methods. Synthesis may be seen as a logical extension of this combined approach. Just as in mixed method primary research, qualitative data can illuminate why a particular policy or management approach works variably and suggest ways of dealing with this, and quantitative studies can indicate its relative effectiveness overall, while both have a contribution to make to understanding the processes shaping implementation of interventions and programmes.

Nature of the guidance for commissioners and researchers

In light of the above considerations, the guidance in this paper is pragmatic, flexible and designed to encourage careful judgement at each step of the review process. Instead of hard-and-fast rules which are not appropriate given the current stage of development of methods for the review of qualitative and quantitative evidence, the emphasis is on encouraging critical thought together with transparency and explicitness about methods. The methods and steps involved in a review incorporating both qualitative and quantitative findings from research cannot be reduced to a fixed series of tasks driven by a single set of standards (e.g. for assessing the quality of evidence). Issues such as whether qualitative and quantitative studies should be synthesized separately or simultaneously can only be resolved in the context of specific projects, informed by accumulated experience.

Stages in a review of qualitative and quantitative evidence

Although the stages of the review process are presented for clarity as if they occur in succession, it is highly likely that the review will be undertaken more iteratively. For instance, scoping the review and mapping the evidence (see below) are likely to take place in parallel with finalizing the specific review questions.

Defining the underlying purpose or aim of the review

It is very important to establish the aim of the review at the outset. Perhaps the most fundamental distinction is between an exercise that is confined to summarizing the evidence (the 'Knowledge Support' approach) and one that includes some or all of the remaining analytical tasks required to reach a decision in the particular policy or management context (the ‘Decision Support’ approach). The latter is inevitably more time, place and value-dependent than the former which tends to strive for greater general applicability at the expense of specific relevance. It is vital that this difference in aim is clarified and agreed at the outset, for example, between the review team and the commissioners of the review, because it affects the choice of approaches to synthesis (see below). In addition, the more a review aims to contribute to a particular decision, the more policy-makers will need either to be members of the review team or, at least, closely involved in the process (e.g. ensuring that the team has access to the necessary ‘soft’ evidence relevant to the particular decision).

The ‘Decision Support’ versus ‘Knowledge Support’ distinction also has implications for the skills of the review team. If the former, it may well be necessary to derive a professional, ‘expert’ or public consensus on an issue where research findings are lacking by applying formal methods of consensus development (e.g. Delphi, nominal group or citizens’ jury techniques). Hammersley distinguishes three distinct aims of ‘Knowledge Support’ reviews:

- aggregative (i.e. focused on the cumulation and generalization of evidence);
- comparative or replicative (i.e. focused on the extent to which different sources of evidence reinforce one another by comparison between sources); and
- developmental (i.e. focused on developing ‘theory’ or explanations).

The first is arguably what many researchers and policy-makers think of when they think of systematic reviews (i.e. using pre-defined methods, including quality standards for deciding which studies to include and focusing exclusively on research). However, the last may be equally important, for example, at an early stage of policy development, generating theories that can inform the development of interventions.

The aim of the review will determine the research questions and is likely to lead to a differing emphasis on qualitative and quantitative data. The more a review aims to contribute directly to a specific decision, the more it will have to include non-research evidence and various methods of modelling and simulation (see below). This will also affect the methodological focus: in ‘Knowledge Support’ avoiding biased findings is central, whereas in ‘Decision Support’ this may be necessary but is not sufficient, since the reviewer must also be explicit about the basis of the judgements inevitably made.

Specifying the review question(s)

Clear review questions are generally essential, whatever the precise approach to be adopted. In a policy or management context, such questions should be developed iteratively at or near the beginning of the review process through discussion between the relevant decision-makers, commissioners and the reviewers as they begin to examine the evidence, and as they learn more about the policy environment in which the review...
is to be used. However, just as in much primary research, additional or amended questions emerge from the process of collecting and analysing the data. In one recent systematic review of both qualitative and quantitative evidence, all the questions were developed from the process of reading and re-reading the articles retrieved, rather than a priori.16

Often the initial questions the policy-maker or manager wants to answer are broad and multi-faceted. For example, a pressing question from a manager or policy-maker might be: ‘What are the costs, benefits and impacts on quality of integrated health care delivery systems compared with non-integrated delivery systems?’ To be feasible as a review, time would need to be spent agreeing a definition of ‘integration’ and perhaps also deciding to focus on particular services to make the scope of the review more manageable. A question such as: ‘Should the government allow private MRI facilities?’ has either to be recast to focus on the costs and benefits of public versus private MRI, if it is to be answered via a ‘Knowledge Support’ review or, if the review is to contribute to a decision as to whether or not to permit private MRI, the reviewers would need access to non-research evidence such as the trade-offs ministers were willing to make between, say, equity, quality, cost and efficiency in the particular country or regional context.

Scoping the review and mapping the evidence

When a review is focusing beyond questions of effectiveness, then it is increasingly recognized as good practice to undertake a preliminary scoping of the potentially relevant literature to ascertain the nature and distribution of relevant studies and other sources, before committing to a final set of questions. This will require some form of search strategy, but in most cases it will be sufficient to restrict this to a search of key electronic databases, unless the reviewers have good reason to believe that the relevant literature is not well covered in such databases. After the review questions have been refined, then a more exhaustive search is likely to be needed – depending on resource availability.

Particularly in areas where there is likely to be a large amount of literature, it will also be necessary to consider the breadth of evidence to be included. In very time- or resource-constrained situations, a ‘review of reviews’, if there is more than one review available, may be the best option, accompanied by a search for studies published since the latest review. The UK Health Development Agency commissioned a number of such reviews, which are well regarded in the health promotion field.17 These may be particularly suitable when there is lack of time to commission an external team and the review is undertaken in-house.

Deciding which forms of evidence to include depends on the aim and questions that the review is to answer. More anecdotal or impressionistic material may be helpful as background (e.g. when formulating the questions for the review), but is less likely to be included systematically in bibliographic databases. If the aim of the project is primarily to develop theory for subsequent testing, then there may be few objections (other than practicality) to the use of a wide range of ‘evidence’ as in ‘realist synthesis’.18 Similarly, if the purpose of the review is to contribute directly to a specific decision in a particular context (rather than simply summarizing the research evidence), then it may be advisable to assemble other kinds of evidence alongside research evidence (e.g. national expert opinion).

Search strategy

The search strategy and subsequent decisions about inclusion and exclusion of studies should be shaped by the review questions, so it is important that the questions are clearly specified. However, no matter how well defined the review questions and scope, searching for studies can be very complex and time-consuming. The broader the review questions are set, the more time consuming and complex the search will be. Ideally, reviewers should make time to develop and test the search strategy to ensure that it is as comprehensive as time allows.

Once the precise review questions have been defined, the preliminary search can be extended. This may include further searching of electronic databases as well as other sources, such as: the internet, conference proceedings, personal contact with researchers in the field, hand searching of specialist journals and scanning of reference lists from studies already identified.

There have been major advances in the development of procedures to search electronic databases for quantitative studies of effectiveness that maximize both comprehensiveness and precision.19,20 However, the relationship between the comprehensiveness and precision of electronic search strategies becomes more problematic when the studies to be identified are more diverse in their focus and design.21 For this reason, reviewers need to be fully conversant with the subject in question to ensure that all relevant search terms are included.22

Searching for qualitative research using electronic databases can be particularly frustrating. There are no equivalents to the Cochrane database of randomized trials for qualitative research. Poor indexing in databases and the diversity of qualitative research make the development of search strategies for identifying qualitative studies difficult.7 Some databases do not index qualitative studies at all (e.g. Medline), while others (such as CINAHL) use a number of methodological indexing terms appropriate to qualitative study designs.23 If the review is going to include a range of study designs, then electronic search terms should not generally involve study design filters. It is also
important to note that similar problems will be encountered in electronic searches for quantitative research other than those on effectiveness.

Despite these difficulties, there are an increasing number of protocols designed to improve the accuracy of electronic searches for qualitative studies; for example, from the Health Care Practice R&D Unit, University of Salford, UK, at http://www.hfsc.salford.ac.uk/hcprdu/projects/qualitative.htm. Similarly, others have adapted protocols designed specifically to search for qualitative studies in systematic reviews. However, research developing and testing different electronic search strategies to identify qualitative studies continues to highlight problems. For example, recent research evaluating the performance of three search strategies for qualitative research on breast-feeding using six electronic bibliographic databases reported that of 7420 records retrieved by the three strategies only 262 were relevant to the review. These findings suggest the need for considerable improvements in the way in which different qualitative research methods are described in journal abstracts and indexed in bibliographic databases.

In most reviews, and particularly those intending to include qualitative research, the reviewers will want to search other sources beside electronic databases. This requires a judgement about how much relevant material may not be covered in these databases and the resources available. In a review of the effectiveness of crime prevention interventions, for example, the authors obtained more than half of their articles from reviewing reference lists and talking to contacts.

Selecting studies and other types of evidence for inclusion

There is no single, simple ‘hierarchy’ of the quality of evidence across all types and methods of research, and across all substantive topics. There is also debate about whether, and, if so, how it is feasible to sample when there are too many sources of evidence to be reviewed. There is the further issue of how much information from a study or other source (i.e. reference, abstract or full text) is necessary before any decision can be made about inclusion or exclusion. There are also debates, particularly in relation to the review of qualitative research about whether a simple include/exclude decision is appropriate at all – the alternative being to take account of quality by allocating different weights to evidence during the synthesis process. This has quite large resource and time implications in many cases.

Quality selection

A key issue in reviewing qualitative and quantitative research is the appraisal of study quality. There is a wide range of views on the selection of studies for reviews. In relation to reviews of exclusively qualitative studies, opinion varies. Some commentators take an extreme view and argue, for example, that quality ‘cannot be determined by following prescribed formulae’ or that it is ‘fruitless to try to set standards for qualitative research per se’. Others accept the need for structured procedures and argue for greater rigour, and standard criteria against which qualitative research may be assessed, but diverge between those who emphasize criteria familiar to quantitative researchers such as inter-observer reliability and construct validity, and those who emphasize more subjective criteria such as authenticity and credibility of accounts. In the health field, the UK NHS Centre for Reviews and Dissemination (CRD) favours a priori structured quality assessment of studies for inclusion in reviews, but recognizes that a vast number of different schema exist for assessing the quality of qualitative studies. Still others argue that while it may be helpful to assess quality, this alone should not lead to exclusion of studies, particularly not in the case of qualitative synthesis (see the section below on qualitative approaches to synthesis). Successful syntheses have knowingly incorporated studies regarded as weak and some synthesis methods counsel that the worth of particular studies may become apparent only during the synthesis process. It may be helpful in these cases to have at least two researchers make the judgement that a study is fit for the particular purpose and perhaps to give its findings a relative quality weighting (see below for discussion of criteria).

Other approaches have few exclusion criteria, but exclude ‘fatally flawed’ studies at a fairly early stage in the review process (but not necessarily before any data extraction and analysis has taken place). The presence of multiple, different checklists and a lack of agreement about quality and exclusion criteria in qualitative syntheses make it impossible to prescribe a single approach to assessing quality and excluding studies from qualitative syntheses. However, this lack of agreed standards indicates that it is prudent to involve more than one researcher in all key analytic tasks.

For quantitative reviews of effectiveness studies, formal hierarchies based on the internal validity of the study design (but ignoring the external validity or generalizability) are often used to set inclusion and exclusion criteria. One of the most sophisticated is the Oxford Centre for Evidence-Based Medicine’s ‘grading system for studies of prevention, diagnosis, prognosis, therapy and harm’. However, the appropriateness of even these hierarchies is disputed, particularly for the quantitative evaluation of more complex interventions. There remains controversy over the assumptions inherent in these quality assessments and how they should be used to influence the overall interpretation of the study findings. There is an acceptance that it is desirable and theoretically possible to use generalizable structured appraisal procedures to minimize error. It is also accepted that clear and transparent approaches for judging ‘good’ research are needed. However, even where clear criteria have been set – for example, the CONSORT statement on randomized controlled
trials, spelling out how they should be conducted and reported – the information required to assess quality is not always available in research reports.

For reviews that attempt to include qualitative and quantitative research evidence, the issue of quality selection is clearly even more problematic. Some structured, generalizable approaches to judging internal validity and reliability in qualitative research are being developed and there are attempts to develop schemes for assessing the quality of quantitative and/or qualitative studies and to distil the common concerns that help unify them (c.f. Mays et al; Spencer et al; Harden et al). Even so, there is a tendency to focus on what might be termed ‘internal validity’ (i.e. faithfulness to the specific context) in these schemes rather than ‘external validity’ (i.e. generalizability), which policy-makers and managers may be as or more interested in. Approaches can be divided into two, depending on whether they propose quantitative criteria (which allow for a cut-off point to be calculated, and thus the exclusion of studies) or qualitative criteria without scores.

According to Murphy et al, validity and relevance represent the underlying criteria which unify seemingly quite distinct approaches. Additionally, as Brannen notes, there are considerable overlaps between general quality criteria used in relation to both qualitative and quantitative research, although there are important differences in how they may be interpreted and applied. The problem facing those wishing to use such criteria to assess qualitative and quantitative research for synthesis is determining which approaches should be used and how the criteria should be applied. Most frameworks fail to specify how judgements should be made on whether or not a standard has been reached. They also typically cannot indicate how much weight should be given to evidence from studies of different quality.

Thus, there is a lack of general guidance on quality criteria applicable to both types of research. The inevitable exclusion of some types of evidence – both qualitative and quantitative – that results from a rigid, hierarchical approach to study quality appraisal remains controversial and may not be beneficial to reviews of qualitative and quantitative evidence. However, reviewers should be aware that some evidence may be of poorer quality (either in its conduct or in its reporting, or both) and they may, in the context of the particular review, need to consider the weight they attach to such evidence and at the very least flag quality issues for the user. Moreover, while this discussion has focused on the assessment of quality in research, the reviewer must address similar issues in relation to any non-research sources of evidence to be included. There is an important point here about the need for careful records to be kept about the appraisal process, ensuring that both the strengths and weaknesses of particular studies and other sources in the context of the particular review are recorded so that this information can be used appropriately in a later stage of the review when the findings are being synthesized and interpreted.

Quantity selection
Sampling is beginning to be discussed in relation to research reviews, though the emphasis in the influential Cochrane Collaboration has been on including all the RCTs in an area, including unpublished studies in order to overcome publication bias. There is clearly a tension between minimizing the potential for publication bias and making reviews manageable. Reviews that involve the transformation of raw data or that include large numbers of studies require greater resources and, where the review question and/or range of evidence is very broad, it may be necessary to sample.

In relation to qualitative studies (and arguably some quantitative or mixed method research, such as process evaluation, and non-research sources of evidence), the main choice is between some form of purposive sampling (e.g. in which studies are sampled according to a set of characteristics such as size, setting and design) or theoretical sampling in which studies are selected for their relevance in testing a particular theory or explanation for a phenomenon. In both cases, the concept of ‘saturation’ familiar to qualitative researchers in which studies are included until new studies cease to provide any fresh data or insights could be used to decide the limits of the review.

Some approaches to synthesis that might be used to integrate qualitative and quantitative evidence in reviews
The synthesis of findings is the heart of any systematic review. There are a number of different approaches currently used for synthesis that have potential to be applied to the synthesis of both qualitative and quantitative evidence. However, practical experience is limited. Most of the methods were developed for primary data analysis rather than synthesis and, of the few approaches to synthesis available, most were developed either for qualitative or quantitative data rather than for synthesizing the two.

A number of superficially distinct approaches exhibit strong underlying similarities. Some approaches present findings in much the same form as those generated by the original researchers, retaining the qualitative or quantitative nature of the original data. Others attempt to convert all data into the same form - either numbers or concepts. For simplicity, we have grouped the different approaches into four basic types.

Narrative approaches
These include traditional narrative reviews of literature which summarize, explain and interpret evidence on a particular topic/question. Such reviews may draw on qualitative and/or quantitative evidence and often include some form of thematic analysis. Findings are
juxtaposed and typically preserved in their original form.

Narrative reviews deal with the findings and interpretations from published studies and other sources in their own terms, without any attempt to transform them into a common metric for analytical purposes. This approach has typically been less systematic and explicit than the other approaches described below, and has been criticized accordingly (though recent narrative reviews have sought to address these criticisms). Notwithstanding its limitations, this approach is flexible, allowing for different types of evidence – qualitative and quantitative, research and non-research – to be reviewed, but generally does not merge findings or generate entirely new theories. This flexibility and ease of handling a very wide range of evidence means that narrative reviews are likely to remain an important tool for policy and management, particularly when time and specialized research skills are at a premium. Narrative reviews can be comprehensive and inclusive since these reviews do not require labour-intensive techniques for data transformation. The other approaches outlined below require some form of data transformation and are thus more likely to have to be undertaken by professional researchers.

Thematic analysis is the most common method adopted within narrative reviews to produce a relatively rudimentary synthesis of findings across the included studies. It seeks to identify and bring together the main, recurrent or most important issues or themes arising from a body of literature. The themes identified will be shaped by the specific review questions.

Increasingly, a range of other methods, such as conceptual mapping and tabular summaries, is being used to introduce a greater degree of synthesis (as opposed to summary) of study findings into narrative reviews. Related to these developments, an important distinction is increasingly being made between the notion of a narrative review and that of a narrative synthesis.

In broad terms, the phrase narrative synthesis refers to a process in which a narrative (as opposed to statistical) approach is used to synthesize evidence extracted from multiple studies. It differs from a narrative review in moving beyond a summary of study findings to attempt a synthesis which can generate new insights or knowledge and be more systematic and transparent.

At least two different approaches to narrative synthesis can be identified. The first of these, which is the subject of ongoing research funded by the UK Economic and Social Research Council (ESRC), adopts the stages of a review described earlier in this paper. The draft guidance produced by this research team argues that a narrative approach to synthesis could be used in three situations: before undertaking a statistical meta-analysis; instead of a statistical meta-analysis where the experimental or quasi-experimental studies included are not sufficiently similar to allow for this; and where the review questions dictate the inclusion of a wide range of different research designs, producing qualitative and/or quantitative findings, and/or non-research evidence, for which no other approach to synthesis is appropriate.

Three main steps in conducting a narrative synthesis are identified:

- developing a preliminary synthesis of the findings of included studies;
- exploring relationships in the findings;
- assessing the robustness of the synthesis produced.

The guidance identifies a number of specific methods that can be used during these processes (including some of the methods described further below), the choice of methods depending on the type of evidence included. As the authors note: ‘the key purpose of this approach to narrative synthesis remains the organization, description, exploration and interpretation of the study findings, and the attempt to find explanations for (and moderators of) those findings’.

Realist synthesis is a second very different approach to narrative synthesis and is discussed in detail elsewhere in this Supplement. In this approach, the primary focus is on the causal mechanisms or ‘theories’ that underlie types of interventions or programmes. This sort of review aims to build explanations across interventions or programmes which share similar underlying ‘theories of change’ as to why they work (or not) for particular groups in particular contexts. Realist synthesis is one way of adding rigour and structure to narrative review, while retaining the ability of the best scholarly narrative reviews to present highly detailed and reasoned arguments about the mechanisms of programme success or failure and about the apparently conflicting results of apparently similar studies.

Narrative synthesis is a relatively new approach and some specific techniques, such as meta-narrative mapping, informed by realist synthesis, have been developed very recently; hence, there are few examples. Box 1 provides one such, misleadingly described in the title simply as a ‘systematic literature review’. Yet, in adopting a systematic, transparent approach to the review and in attempting to use synthesis to generate new insight, it is a form of narrative synthesis rather than a narrative review. This review provides an account of how, why and in what sequence, a field of research has unfolded, enabling the reader to see how explanations (theories) and empirical findings have intertwined and changed one another through time. The guidance on narrative synthesis funded by the ESRC is currently being applied to practical review questions; so more worked examples of narrative synthesis will be available in the future.
matrices or tables to allow comparison of studies. Examples of this approach include meta-ethnography and qualitative cross-case analysis.

Meta-ethnography is a way of using qualitative research techniques to synthesize multiple qualitative research reports.33 It seeks to do more than simply collect and review a series of accounts (as in a narrative review) and instead aims at a novel synthesis, which develops a new theory to explain the range of research findings encountered. It is a way of re-analysing and comparing the texts of published studies (rather than the original data of each) to produce a new interpretation. There is nothing, in principle, to stop its use for the synthesis of quantitative and qualitative research findings, though it has not yet been used in this way. A particular value for policy-makers of this approach is that it could help explain seemingly divergent quantitative and qualitative study findings.

Primary research using a case study approach may be based on single or multiple cases, and can include a mixture of qualitative and quantitative evidence.44 Qualitative approaches to cross-case analysis, which typically compare findings from different sources using some sort of matrix or text-table format to juxtapose data, provide another potential method for qualitative-quantitative synthesis (see the Appendix of the web version of this report published by the funders at http://www.chsrf.ca/funding_opportunities/commissioned_research/projects/msynth_e.php or at http://www.sdo.lshtm.ac.uk/researchmethods.htm for further details on this and the other approaches described here).

A demonstration of the meta-ethnographic approach is provided by Britten et al.45 They synthesized the findings of four qualitative studies relevant to the question, ‘How do the perceived meanings of medicines affect patients’ medicine-taking behaviours and communication with health professionals?’ The authors identified the common and recurring concepts in each paper (preserving the interpretations offered in the original papers) and displayed these in a matrix. This allowed a systematic process of comparison and ‘translation of the studies into one another’ (i.e. showing how a concept in one paper was encompassed in interpretations offered by the other papers). In turn, this led to the development of new interpretations (referred to as ‘third order interpretations’) and a line of argument that there were two quite distinct forms of medicine-taking, one adherent to instructions and the other based on self-regulation (see Table 1). Meta-ethnography has been used to synthesize larger numbers of studies,32 but there are still relatively few examples of its use in health research.

Quantitative approaches

These convert all data into quantitative (i.e. numerical) form. This entails transforming qualitative findings into numbers by identifying themes which can then become variables that can be quantified either as frequency counts or in binary form. The quantitative case survey method is a formal process for systematically coding data when there is a large enough number of independently conducted qualitative studies to allow for quantitative analysis.46 A set of structured questions is used to extract data from each
case study, which is then treated as an observation within a single data-set. Data are then converted to quantitative form for statistical analysis, which can include meta-analysis (see below) or other quantitative techniques. Box 2 gives an example taken from Yin (2003). Case survey is an explicit way of turning qualitative studies into quantitative data for analysis, thereby allowing an integrated qualitative-quantitative synthesis to be undertaken. If there are too few studies for quantitative analysis, the method can be used to create text charts of themes which are similar to the qualitative cross-case analysis described above.

Content analysis is a systematic technique for categorizing data into themes, developed for primary research and involving a wide variety of mainly textual information. It can be used in the synthesis of findings from multiple studies to count how often each category or theme occurs in order to identify dominant findings and make generalizations. Stemberg defines content analysis as ‘a systematic, replicable technique for compressing many words of text into fewer content categories based on explicit rules of coding.’ Content analysis is essentially a quantitative method since all the data are eventually converted into frequencies, though qualitative skills and underlying theory may be needed to identify and characterize the categories into which findings are to be fitted (thematic analysis is its qualitative counterpart – see above).

Bayesian meta-analysis and decision analysis

In this group of techniques, all data are converted into quantitative form and pooled for analysis and modeling (more details are given in the Appendix to the web version of this report [see previous page for link]). These methods are designed primarily for the Decision Support function and, accordingly, allow the incorporation of a wide range of evidence, including findings from qualitative research and subjective judgements (e.g. public or decision-makers’ preferences and the trade-offs they are willing to make between policy goals derived from consensus methods) in a quantitative synthesis.

The Bayesian approach has been defined as, ‘the explicit quantitative use of external evidence in the design, monitoring, analysis, interpretation and reporting of a study’ and can be applied to synthesis. Under this approach all qualitative evidence has to be converted into quantitative form. This approach has the attraction of allowing potentially any form of relevant evidence to be brought to bear on a decision. As a result, sources such as expert judgement and public and/or stakeholder opinion and preferences can be used as inputs to meta-analysis, quantitative modeling and simulations of effectiveness and cost-effectiveness. The main drawback concerns the feasibility of converting complex, implicit judgements into specific

---

**Box 2** An example of the case survey method – synthesis of evaluations of community crime prevention

Eleven evaluations previously conducted by different investigators using a variety of mostly quantitative methods, of 11 different forms of crime prevention programme, were dissected and compared to identify generalizable features of successful programmes which were independent of any specific form of intervention. All 11 evaluations had been deliberately chosen because they showed positive results despite each focusing on a different type of programme.

The findings would enable policy-makers and programme managers to specify the minimum essential characteristics that all future programmes should exhibit to increase their likelihood of success.
weights and probabilities, though there are methods available, for example, to convert qualitative probabilistic judgements such as ‘x is extremely possible’ or ‘x is slightly possible’ into numerical weightings. Box 3 gives an example of an innovative Bayesian synthesis which draws on findings from qualitative and quantitative research to identify and assess the relative importance of a range of factors potentially affecting the uptake of childhood immunizations. The study shows the benefits for policy and management of trying to include both qualitative and quantitative evidence, since there were potentially important factors that would have been omitted had one or the other source of evidence been relied upon.

Choosing an approach
Choice of a particular approach(es) will depend both on the underlying aim (i.e. review of the evidence base or direct contribution to a decision – reaching a ‘verdict’), the specific questions to be addressed, the nature and balance of evidence available (e.g. whether most of the research is qualitative or quantitative), and the stage that policy development has reached.

At an early stage in policy development, a review designed to develop ‘theory’ (i.e. an explanation for a phenomenon) might be most helpful. In this case, a narrative synthesis could be appropriate. At a later stage, the focus of any review might be more on the different cost-effectiveness of different interventions and of the same intervention in different settings in which case a combination of cost-effectiveness modelling and meta-ethnography could be helpful.

Table 2 attempts to provide some guidance on which approaches are likely to be suitable for which sorts of review aims and questions.

A review may comprise a combination of approaches to synthesis, so that a wide enough range of evidence can be brought to bear on a policy or management question. For example, a review using thematic analysis could be followed by a meta-ethnography to develop a ‘line of argument’ synthesis to arrive at some higher-order interpretation. A quantitative approach or a narrative review might generate additional questions that could be pursued via either a qualitative synthesis or a data-pooling method such as Bayesian meta-analysis. We are not aware of any such examples to date.

An example of how a combination of approaches can be used is provided by the work of a team at the University of London, Institute of Education. This team has conducted a series of reviews that synthesize a range of qualitative and quantitative evidence to

---

Box 3 An example of a Bayesian synthesis of qualitative and quantitative research evidence

**Objectives**
To identify factors potentially affecting uptake of childhood immunization in developed countries.
To assess the probable effect of these factors on levels of uptake.

**Methods**
1. Generation of a prior distribution of factors from qualitative findings
   A prior distribution of factors likely to affect the uptake of childhood immunization was derived from the subjective views of the researchers (n = 5). This was revised after the researchers had read all relevant qualitative studies and extracted factors from each study, ranked them in order of importance and then revised their first opinions on the factors and their ranking. Rankings of each factor given by each researcher were combined to yield an overall probability that the factor would be important in determining uptake of immunization.
2. Extraction of quantitative data
   Data extracted from quantitative studies according to categories (factors) generated by analysis of qualitative studies, with extra categories (factors) generated as needed if not included in the qualitative studies.
3. Statistical analysis
   The prior probability of a factor being important was combined with the quantitative evidence to form a posterior probability that each factor (e.g. lay health beliefs) was important in affecting uptake of immunization. Regression models for the odds of uptake were constructed using relevant data from the quantitative studies for each factor in turn.

**Results**
Qualitative and quantitative studies identified some common factors, but others were reported only in qualitative or quantitative studies. For some factors, addition of quantitative data substantially modified prior probabilities.

**Features**
Inclusion of qualitative and quantitative studies allowed a wider range of potential factors affecting the uptake of immunization to be identified and assigned a probability of importance
Qualitative and quantitative studies contributed distinctly different evidence relevant to improving take-up (e.g. quantitative studies highlighted the effect of socioeconomic factors and qualitative studies showed the importance of lay health beliefs).
<table>
<thead>
<tr>
<th>Review aim and/or policy/management question</th>
<th>Relevant types of evidence (if available)</th>
<th>Likely approach(es) to synthesis</th>
<th>Strengths of the approach</th>
<th>Weaknesses of the approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>'Knowledge support'</td>
<td>All types, but mostly research-based because of focus on what the research evidence says</td>
<td>All bar decision analysis, etc.</td>
<td>Focuses on the specifics of a particular decision in a particular context</td>
<td>Does not directly help with a specific decision in a particular context</td>
</tr>
<tr>
<td>'Decision support'</td>
<td>All types, including research and non-research (i.e. need to know evidence, values and preferences of stakeholders and decision-makers)</td>
<td>Bayesian meta-analysis, decision analysis, modelling and simulation of various types and possibly narrative synthesis</td>
<td>Has to be modified to be relevant to another context; utility depends on its being used by decision-makers, not generalizable</td>
<td></td>
</tr>
<tr>
<td>Is this a problem?</td>
<td>All types including research and non-research (e.g. public &amp; stakeholder views, opinion polls, focus groups), qualitative and quantitative</td>
<td>Narrative synthesis (e.g. meta-narrative mapping); or, for qualitative studies, meta-ethnography</td>
<td>Narrative synthesis is flexible, relatively easy to understand and applicable to a range of situations and sources of evidence; meta-ethnography is less flexible in that it has yet to be used for other than qualitative synthesis</td>
<td>Have to work hard to make sure methods and judgements are explicit, free of bias and replicable; defining something as a ‘problem’ is value-laden</td>
</tr>
<tr>
<td>How big is the problem? Which groups does it affect?</td>
<td>Quantitative research and routine admin. data. Qualitative data on subjective impact</td>
<td>Quantitative synthesis plus meta-ethnography of qualitative studies</td>
<td>Meta-ethnography is labour intensive, requires considerable qualitative research experience</td>
<td></td>
</tr>
<tr>
<td>Is it changing over time?</td>
<td>Quantitative research and routine admin. data</td>
<td>No need for qualitative–quantitative review, straightforward quantitative synthesis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>What can be done about it (what may work)?</td>
<td>Mostly quantitative research on effectiveness and cost-effectiveness</td>
<td>Meta-analysis of intervention studies</td>
<td>Meta-analysis well developed for effectiveness and reasonably well developed for cost-effectiveness data</td>
<td></td>
</tr>
<tr>
<td>What works, for whom, in which circumstances?</td>
<td>Wide range of research and non-research data</td>
<td>Realist synthesis; narrative review; case survey</td>
<td>Helps with understanding mechanisms underlying interventions (how they work) – enlightenment role of evidence</td>
<td>Will not necessarily produce specific answers to particular decision needs</td>
</tr>
<tr>
<td>Will intervention/policy x work here with what cost and benefit consequences?</td>
<td>Cost-effectiveness data from research; modelling related to specific circumstances including non-research data</td>
<td>Bayesian meta-analysis and cost-effectiveness modelling; decision analysis</td>
<td>Makes research evidence relevant to specific circumstances of a particular decision</td>
<td>Dependent on validity of expert opinion where research is lacking and on specific value trade-offs of decision-makers; decision-makers may be reluctant to follow 'verdict' of the analysis; Bayesian meta-analysis can be hard to explain</td>
</tr>
<tr>
<td>How acceptable will intervention/policy x be?</td>
<td>Largely qualitative research and non-research data (e.g. focus groups, opinion polls, stakeholder analysis)</td>
<td>Qualitative synthesis (e.g. meta-ethnography of research, qualitative cross-case analysis, etc.)</td>
<td>Essential information for policy-makers and managers even though tricky to interpret</td>
<td>'Softness' and/or transitory nature of opinions and views</td>
</tr>
</tbody>
</table>
answer the overarching question: ‘what is known about the barriers to, and facilitators of, people’s mental health, physical activity, smoking and healthy eating?’

The reviews comprise three different approaches to synthesis:

1. a conventional synthesis of quantitative effectiveness studies using statistical meta-analysis (pooled data and calculation of overall effect sizes), where appropriate;
2. a qualitative synthesis of studies of people’s perceptions and experiences in which studies are interrogated, re-analysed and combined to produce an overarching thematic analysis or ‘theory’ as to which kinds of interventions might work to promote participants’ health. The interpretive framework for the thematic analysis is driven by the review objectives (e.g. identifying barriers and facilitating factors to specific health behaviours); and
3. a cross-study synthesis which combines the ‘theory’ of synthesis 2 to interpret, add context and refine the conclusions of the quantitative synthesis 1. This can involve further meta-analyses to test hypotheses about factors influencing the effectiveness of interventions.

The principles underlying this combined approach are: to preserve the unique contribution of qualitative and quantitative evidence while also providing a way for each type of evidence to help interpret the other, in order to form a more comprehensive and useful

**Figure 1** Steps in a qualitative and quantitative narrative review. Reproduced from *J Epidemiol Community Health* 2004;58:794–800, with permission from the BMJ Publishing Group.
answer to the review questions, and to recognize explicitly that different types of questions require different sorts of evidence (e.g. qualitative evidence cannot be used to assess the average balance of benefit and harm from an intervention and that the quantitative evidence relevant to this assessment cannot describe people’s experiences of the intervention).

For example, the team’s review of smoking cessation interventions for pregnant women incorporated qualitative data and, distinctively, new evidence from consultations with key stakeholders in order to be able to juxtapose women’s views with a systematic review of quantitative evidence from trials of the effectiveness of interventions. As a result, both process and outcome measures could be taken into account. The qualitative material revealed outcomes important to women, not addressed in the trials (e.g. the impact of smoking on method of delivery, breast-feeding and maternal well-being), which should be taken into account in the design of policies and future interventions to reduce smoking in this group.

Figure 1 shows the steps undertaken in another review by the same team focused on identifying barriers to, and facilitators associated with, health and health-related behaviour among young people. The team combined qualitative and quantitative evidence in a mosaic of findings that preserved the unique contribution of each, rather than attempting to pool all the available data. In this way, qualitative studies of young people’s views were used to assess the interventions found to be effective in the quantitative evaluative research and, as a result, it was possible to identify which properly evaluated interventions were both effective and acceptable to young people.

What is a good-quality review?

Rigorous systematic reviews of qualitative and quantitative evidence are still at an early stage of development and so it is not possible to provide definitive quality criteria. On the other hand, it is possible to identify a set of questions which could usefully be asked, either of a planned or completed review designed to be used for policy or management purposes:

- is the aim of the review clear (e.g. ‘knowledge support’ (enlightenment, comprehension, insight, etc.) or ‘decision support’ (which goes beyond synthesizing the evidence to provide a policy solution)?
- are the review questions relevant to the concerns of managers and/or policy-makers in the particular setting?
- are the methods explicitly and comprehensibly described (such that another team could repeat the review using the same methods)?
- is each step in the method clearly justified (are there signs of the use of informed judgement in deciding how to tackle the review)?
- is there an appropriate ‘fit’ between the questions posed and the type(s) and sources of evidence brought together for the review (does the review include a sufficiently wide and relevant scope of evidence)?
- did the review team include the range of disciplines, skills, subject area knowledge and experience necessary to answer the question(s)?
- do the reviewers show that they understand the context in which the review’s conclusions will be used for policy and/or management decision-making?
- where judgements are made on the basis of the evidence brought together, is the reasoning underpinning these judgements made clear, so that it can be discussed, if necessary, and revised, if found wanting?

Conclusion

Policy-makers and managers increasingly require access to high-quality evidence syntheses that include research and non-research evidence, and both qualitative and quantitative research findings. There is no single, agreed framework for synthesizing such diverse forms of evidence and many of the approaches potentially applicable to such an endeavour were devised for either qualitative or quantitative synthesis and/or for analysing primary data. However, there is scope to develop and adapt the range of approaches discussed above.

At present, narrative review in some form is likely to remain the basis of most attempts to review and synthesize mixed sources of evidence. However, there are signs that narrative reviews are becoming more transparent and systematic and striving for narrative synthesis (see above) rather than simply juxtaposing the findings of studies. Other more integrative methods are still evolving and/or are untested, and may only be relevant to part of the evidence in a review for policy or management purposes. Finally, it is worth noting that, despite the focus of this paper, a synthesis of qualitative and quantitative research will not invariably be needed, given that each normally tackles different sorts of questions.

Acknowledgements

We gratefully acknowledge the contribution made to this paper by the peer reviewers appointed by the funders (Linda Bilheimer, Mary Dixon-Woods and Ron Stamp), and by the managers and policymakers who acted as discussants at the project workshop in London, 16-17 May 2004. Thanks too to Angela Harden for discussion of the review methods developed by the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) at the University of London, Institute of Education. Without the help and acute observations of all these people, this would have been a poorer product. Fortunately, they can take no responsibility for the use we made of their helpful observations. We are pleased to acknowledge the financial support of the Canadian Health Services Research Foundation in association with the English NHS R&D Service Delivery and Organisation Programme.
References


50 Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. Bayesian methods in health technology assessment: a review. Health Technol Assess 2000;4