School for Social Care Research

NHS
National Institute for
Health Research

Systematic reviews in social care and social work research

Deborah Rutter

Methods Review 13

Improving the evidence base for
adult social care practice

The School for Social Care Research

The School for Social Care Research is a partnership between the London School of Economics and Political Science, King's College London and the Universities of Kent, Manchester and York, and is part of the National Institute for Health Research (NIHR) http://www.nihr.ac.uk/.

The School was set up by the NIHR to develop and improve the evidence base for adult social care practice in England. It conducts and commissions high-quality research.

About the author

Deborah Rutter is a Senior Research Analyst at the Social Care Institute for Excellence (SCIE), contributing to the production of guides, rapid systematic reviews and research briefings. SCIE supports the development of the evidence base for social care practice by identifying and analysing knowledge from different sources about what works in social care, and translating that knowledge into practical resources, learning materials and services, including training and consultancy. SCIE uses methods of systematic review and synthesis, tailoring them to different types of research question and available evidence, and incorporating co-production strategies with users and providers in the sector. Previously, Deborah was a Research Fellow at Imperial College Medical School. Her research interests and experience include mental health and personality disorder services, substance misuse, user involvement, end of life care, qualitative and ethnographic methods and research ethics.

NIHR School for Social Care Research
London School of Economics and Political Science
Houghton Street
London
WC2A 2AE
Email: sscr@lse.ac.uk
Tel: +44 (0)20 7955 6238
Website: www.sscr.nihr.ac.uk

ISBN 978-0-85328-471-0

ABSTRACT

This review aims to introduce the reader to the concept and methods of systematic reviewing of research evidence, and to some of the debates that consider the use of systematic reviews for knowledge production in, and dissemination to, social care settings. It outlines the stages of systematic reviewing, including topic selection, drawing up and piloting of inclusion criteria, searching, quality appraisal, data extraction and synthesis, and reporting. The review highlights some of the methodological challenges arising from the use of systematic reviews within social care and social work value and resource systems (e.g. involving stakeholders and end users in the work; setting review parameters to fit available resources; the limitations of primary studies in social care settings), and considers the different types of review questions and research evidence with which reviewers in social care may engage. Some limitations of systematic reviews, including the difficulty of conveying nuanced conclusions to policymakers, are described. Since this review was written, there have been further valuable publications in this field, highlighting the importance of using reviews of evidence to underpin practice within social care at a time of funding cuts.

RECOMMENDATIONS FOR RESEARCH ON ADULT SOCIAL CARE PRACTICE

- More research in social care should be funded. The present system of funding health and social care predominantly through NHS portals does not reflect the need for a stronger evidence base in social care. This evidence base is needed to support systematic reviews, and systematic reviews are needed to support practice.
- The commissioning of systematic reviews should be informed by early exploration of the literature. If available primary research cannot support a review, funding may be better spent on commissioning primary research.
- Researchers in social care should be more ambitious in employing a range of research designs. In particular, they should consider the feasibility of controlled studies to improve the evidence base for the benefit of the sector. This may be difficult and costly to achieve in settings such as residential care, and funding should reflect the apparent obstacles.
- Studies of effectiveness should be supported or supplemented by qualitative research on the implementation and acceptability of approaches from different perspectives, enabling a range of research questions to be approached through systematic research and review techniques.
- In an age of relative austerity, research providers are often required to find faster and smarter ways to interrogate the evidence base for social care. Rapid review techniques, transparently reported, are increasingly becoming the norm. Some methodological work on the relative benefits and shortcomings of comprehensive vs rapid reviews would be useful.

- Social care commissioners, social work students, and social care providers and users are often not well prepared to understand the implications and shortcomings of findings from research reviews. Formal and informal training promoting ‘research-mindedness’ would be useful to the sector.

KEYWORDS

Systematic reviews, evidence-based practice, RCTs, quantitative vs qualitative designs

ACKNOWLEDGEMENTS

The groundwork for this paper was provided by the practice and insights of SCIE’s Research and Reviews, and Searching and Scoping, Teams. As will be clear, the paper has attempted to reflect the work of the many research practitioners who have contributed a wealth of expertise to this topic. A special acknowledgement is due to Prof. Mike Fisher for his support in this, and many other, methodological excursions.

CONTENTS

INTRODUCTION TO SYSTEMATIC REVIEWS

Why do we need systematic reviews of evidence?

Knowledge is a highly contested field. We can probably all point to personal experiences of care as evidence. My grandfather, for example, never accepted that cigarettes cause cancer, because his ‘evidence’ came from his own long, smoke-filled life and eventual death from other causes. His interpretation of the damage wrought by nicotine was not informed by public health programmes. People receiving social care support might all point to various experiences as evidence of what has worked. Yet using such personal experiences as the basis for deciding more general issues of social policy and practice would obviously be highly contentious and risky. Where public money is to be spent on education and services to improve health and wellbeing, we need methods of arriving at a consensus on what reliable sources of evidence there are, and what that the evidence is telling us.

Systematic reviews are a means of building bodies of evidence about a research topic or question from unrelated research studies. They follow a transparent path from the evidence toward a defensible conclusion, seeking to minimise as far as possible personal and methodological bias. Systematic reviews also offer a summary, which is valuable in a world where those that need to know cannot possibly access and read all the evidence. Hence, ‘The purpose of a systematic review is to sum up the best available research on a specific question’ (www.campbellcollaboration.org). More specifically, this may include:

- to make sense of an information explosion by bringing together and exploring gaps and weaknesses in the knowledge base;
- to influence decision making or to legitimise action;
- to generate new insights and understanding, for example by confirming or modifying theory (Braye and Preston-Shoot 2007).

What is a systematic review of evidence?

There is strong agreement on the definition of a systematic review of evidence, although there is some debate about how these are best carried out and what should be included.

A systematic review is a ‘review that strives to comprehensively identify, appraise, and synthesize all the relevant studies on a given topic. Systematic reviews are often used to test just a single hypothesis, or a series of related hypotheses.’ (Petticrew and Roberts 2006). A review is systematic because it ‘uses explicit and systematic methods to identify, appraise and summarise the literature according to predetermined criteria. If this description [of the criteria and methods of applying them] is not present [e.g. in the protocol], it is not possible to make a thorough evaluation of the quality of the review.’ (National Institute for Health and Clinical Excellence (NICE) 2009).

Box 1: Definition of a systematic review

A systematic review uses transparent procedures to find, evaluate and synthesise the results of relevant research. Procedures are explicitly defined in advance, in order to ensure that the exercise is transparent and can be replicated. This practice is also designed to minimise bias.

Studies included in a review are screened for quality, so that the findings of a large number of studies can be combined. Peer review is a key part of the process; qualified independent researchers control the author's methods and results.

A systematic review must have:

- Clear inclusion/ exclusion criteria
- An explicit search strategy
- Systematic coding and analysis of included studies
- Meta-analysis (where possible).

(www.campbellcollaboration.org)

The transparency of the methods used in a systematic review is important: it should be possible for other researchers to replicate the entire review, although this would hardly be worthwhile unless there were new evidence to be incorporated.

There are then a number of features that make a review systematic. These include:

- **The comprehensive nature of the search for evidence**, often international (though often constrained by language; see (Grégoire *et al.* 1995; Moher *et al.* 1996);
- **Clarity and transparency of methods and process of review**, laid out within a protocol. 'The protocol specifies the plan which the review will follow to identify, appraise and collate evidence' (NHS Centre for Reviews and Dissemination 2001). The protocol is usually itself peer reviewed;
- **Effort to limit or take account of bias** in the research on a topic. 'Systematic review (is) the application of strategies that limit bias in the assembly, critical appraisal, and synthesis of all relevant studies on a specific topic' (Chalmers *et al.* 2002);
- **Synthesis of accumulated research** on a topic. A systematic review offers 'a synthesis of results of existing evaluations and research projects produced with the purpose of clarifying [a question such as] whether a given intervention works' (Hansen and Rieper 2009);
- **A systematic review can be updated**, because the method is transparently reported. This is important, given the time invested in some reviews, and the developing evidence base.

Hammersley (2001) has been critical of the dominant approach to systematic reviews. His criticisms include that they are very proceduralised, and how they tend to privilege certain kinds of knowledge and, hence, how they exclude other knowledge. Taylor *et al.* (2007) offer a framework intended to be more inclusive in the appraising of different research evidence for consideration in a systematic review.

Systematic reviews are more widely used in health sciences than they are in social care, and one focus of the literature referred to in this paper is the adaptation of the methods to suit social science applications and social science data. This account will give an outline of the methods of review, and flag up some of the debates around their application to social science topics. It is probably apt here to make the following disclaimer:

This paper is not based on systematic review methods – it does not address a review question and has no pre-determined protocol for identifying, assessing and synthesising material. It is instead a personal overview, using literature sources valued by, and available to, the author (largely via the Social Care Institute for Excellence).

A brief outline of systematic review methods

A systematic review brings together all the available evidence, identifies commonalties and differences, identifies gaps in the evidence base, and should allow synthesis of findings. Such a synthesis is particularly useful where primary studies have similar characteristics in method, outcome measures and, of course, the research question they aim to answer. The following abbreviated outline lays out the stages of a review, while each aspect is explained in greater detail in the section 'Methods of systematic reviews'.

The classic procedure for a systematic review requires reviewers to:

1. Clearly define the question the review is setting out to answer
2. Determine the types of studies needed to answer this question(s) (i.e. draw up inclusion criteria)
3. Carry out a comprehensive search to locate studies for inclusion
4. Screen the results of that search against inclusion criteria
5. Critically appraise the quality of these included studies
6. Extract data from each proficient study to aid analysis and draw up evidence tables
7. Synthesise the studies and assess heterogeneity
8. Report and disseminate

(Adapted from Petticrew and Roberts (2006): item 6 inserted)

The following section considers the background to, and need for, systematic reviews in public policy and in social care.

EVIDENCE AND SOCIAL CARE RESEARCH

The role of systematic reviews in evidence based policy and practice

Science works on a cumulative approach to knowledge and understanding, and although evidence synthesis in reviews has been a feature of scientific practice for centuries, it is only fairly recently, and particularly with the founding of the Cochrane and Campbell Collaborations (see below for more information on these), that reviews have achieved the prominence they have today. Interestingly, the production of reviews was once derided as ‘parasitic recycling’ of the work of others (Chalmers *et al.* 2002).

One of the benefits of using reviews to address questions concerning best practice is that different studies take place in different contexts, so there are good opportunities to test the effectiveness or suitability of an approach in different settings. This is important when we think about the generalisability of approaches to the wide range of social care settings. Comparing and exploring the variation of results from different studies included in reviews can tell us a lot about universal effectiveness.

In health technology studies, what works for whom is traditionally determined by applying systematic review methods to controlled studies of pharmaceutical products and defined health interventions or technologies. Within social care, this type of raw material from controlled studies for use in reviews is rather less common and less well resourced (Marsh and Fisher 2005), and some social scientists do not find systematic reviews based on experimental methods fit for their purposes or their research topics. However, systematic reviews can and do draw together different types of evidence, and have the potential to inform policy and practice, to establish the gaps in the evidence base and to identify harmful practice beyond reasonable doubt, so they are a valuable tool for policymakers, practitioners, managers and commissioners, as well as for researchers. The rhetoric of ‘evidence-based policy and practice’ is now part of political discourse, if not always practice, and systematic reviews can be powerful sources of evidence. ‘We need to be able to rely on social science and social scientists to tell us what works and why and what types of policy initiative are likely to be most effective’ (David Blunkett (then Secretary of State for Education) quoted in Boaz *et al.* (2002)).

However, politicians have other drivers for generating policy, such as public pressure, and the role of evidence in formulating policy is clearly partial and often unclear. The uncertain influence of reviews on policy may reflect the lack of primary research available to answer complex questions, as well as the costs and time delays inherent in the commissioning of reviews.

Alongside political acceptance of the need for evidence to support existing and innovative practice, the evidence movement, an informal network of evidence-producing organisations, has grown to produce and disseminate systematic reviews (Hansen and Rieper 2009). The Cochrane Collaboration was established in 1993 to, as they phrased it, rescue the UK NHS from ‘collective ignorance about the effects of many common aspects of healthcare' (Chalmers *et al.* 2002); a sibling organisation, the Campbell Collaboration, was established in 2000 to produce reviews in the fields of social welfare, criminal justice and education. The funding of EPPI-Centre in 1993 at London University's Institute of Education (Oakley *et al.* 2005) has contributed much to the methodology of reviews in social sciences, as well as to the evidence base for education. The What Works Clearinghouse was established in 2002 by the US Department of Education. Government funds have also supported the Centre for Reviews and Dissemination at York, where the DARE database of reviews is kept. Both NICE and the Social Care Institute for Excellence (SCIE) (Fisher 2002) utilise these resources, and commission systematic reviews to support the publication of guidance in health and social care.

It remains the case that for many reasons civil servants, commissioners and practitioners continue to struggle to apply the subtle messages of research to practice (Bero and Jadad 1997). Furthermore, as Boaz and Pawson (2005) note, researchers to date may not have been the best at converting review findings into implications for practice. The transfer of knowledge into practice is in itself a huge topic which cannot be addressed here: Mullen *et al.* (2005) consider the challenges to evidence based social work.

Although social scientists are increasingly engaging with reviews and review methodology, the inadequacy of research funding for social care and social work practice (Reilly *et al.* 2008) has led to a context where evidence in these fields may be 'hard to come by, of questionable quality and uncertain relevance' (Gowman and Coote, quoted in Boaz *et al.* (2002)), a situation not confined to UK social services (Olsson 2007). The following section comments on the sources of evidence which are the building blocks of systematic reviews in health and social care, and the general calibre of evidence available to the reviewer.

Introduction to social care research evidence

Systematic reviews have been developed, promoted and disseminated by Cochrane Collaboration and the Campbell Collaboration (respectively known as C1 and C2) as the best level of evidence of effectiveness for defined interventions (Davies and Boruch 2001; Chalmers *et al.* 2002). Questions tackled by systematic reviews of this pedigree commonly relate to questions of effectiveness or cost-effectiveness of an intervention, drawing on Randomised Controlled Trials (RCTs)[1], which use quantitative measures of outcomes, allowing the size of effects to be calculated and compared to those arising from a control group who did not have the intervention. The different studies that meet the review criteria are synthesised (through a meta-analysis) to give an overall effect size for a group of studies. The appropriateness of this method has been widely criticised for a number of

---

[1] A randomised controlled trial is a research method where groups of service users are assigned by chance, usually in equal proportions, either to receive a service, intervention or approach, or to receive the standard treatment. What happens to each group is then assessed, and the assumption can be made that differences relate to the different treatments received. In practice, other factors may bias the findings and confound the effect of the new service, so studies are critically reviewed for these factors.

reasons, some of which are explored below (see section 'Methodological issues and challenges to the methodology of systematic reviews').

The use of RCT methodology in social care and social work research is more unusual than it is in healthcare, and many argue that there are good, pragmatic reasons why this is the case. Social care is often less technically driven, and successful outcomes may depend on the quality of human interaction, which is difficult to standardise and measure. Social care research has historically been underfunded, and RCTs are expensive. Social care and social work is delivered in highly variable contexts and is affected by implementation factors which must be relevant to outcomes: yet RCTs focus on inputs and outcomes and minimise the significance of factors and processes in between these points. In social care, it may be difficult to provide a control group – a group of participants who are similar in every respect except that they are not receiving the treatment in question. It may be that service users will not agree to take part in an RCT if they think the new service is a definite improvement; or, that the new approach requires such a high degree of co-production that the preferences of people for one approach over another must be taken into account, thereby foiling the random assignment aspect of the method. Reviewers seeking some kind of control or comparison group may then look for before-and-after studies, where everyone receives the new service, and outcome measures are recorded both before and after a new approach is introduced. Case control or matched studies may also be ways of providing a comparative element. A key issue in applying the methodology of systematic review to the social sciences is then to decide which types of study should be included in the pool of relevant research to be plumbed and how reliable these study methods are.

The topic of the review and the questions it addresses are central to the inclusion of different study types or designs. Many outcomes of social care, health and education interventions are not apparent until the receiving child becomes an adult. Cohort studies may be the best type of evidence to include in a study of health and educational outcomes of early years support for young children (Coghlan *et al.* 2009), or the impact of cannabis smoking in young people on their mental health as adults (Kuepper *et al.* 2011). If the review question seeks to understand the incidence of a social need in a population, then correlates studies, including the use of large-scale datasets, may be the key source, while interrupted time series studies will show the impact of new legislation by taking measures at different time points. There is no limit on the type of study that can be included in a systematic review, though it is sensible specifically to exclude papers that have no basis in research, such as opinion pieces or articles in professional journals.

This paper also argues, and assumes, that systematic reviews in the social sciences will include qualitative studies if they can contribute to answering the review question or questions. A review may involve a set of related questions, and the type of material included may vary for each question: controlled studies may tell us 'what works best', and qualitative research may provide valuable material on issues of implementation and stakeholder views. From a nursing perspective, Lloyd Jones (2004) points out that the inclusion of qualitative research papers in a systematic review adds substantially to the time required to conduct a more traditional systematic review of randomised controlled trials, and a substantially larger proportion of papers, including those without abstracts, may have to be retrieved and read in full text before inclusion is certain. Indeed, the time cost of reviewing is a probable and rational motive for relying on limited approaches and research designs, but the type of evidence included in a review should depend primarily on the nature of the questions the review seeks to address.

Systematic reviews are not then synonymous with meta-analysis of randomised or controlled trials. The key features of systematic reviews are the systematic and transparent nature of the review process and the appraisal of included research studies (of all types included) for sources of bias.

Two examples of systematic reviews relating to social care are:

- Mason *et al.* (2007) on models of community-based respite care for frail older people and their carers;
- Parker *et al.* (2010) on integrated models of care for people with long-term neurological conditions.

Other examples are given in the table towards the end of this paper.

METHODS OF SYSTEMATIC REVIEWS

The Cochrane Handbook for Systematic Reviews (Higgins and Green 2006; Littel *et al.* 2008), offer detailed accounts of the methods of systematic reviewing, which cannot be replicated here. A short journal article on the stages in a systematic review can be found in Wade *et al.* (2006). The methods outlined in these resources are entirely suited to the social sciences and to social care, although as we have seen, the types of evidence available in other sectors may steer the review questions and methods. Systematic reviews in the social sciences are covered specifically in Petticrew and Roberts (2006), and SCIE has produced guidance for commissioned reviewers (Rutter *et al.* 2010). The following brief account of review methodology derives from the stages identified in Petticrew and Roberts (2006) (see section 'A brief outline of systematic review methods' above).

Defining the question the review is setting out to answer

Formulating one or more review questions is often a consultative activity. Reviews use scarce resources, time and expertise. At SCIE, reviews are commissioned to support programmes of work and may be supported by initial scoping or mapping. Some initial investigation of the available evidence is also useful at this stage: there is no point in planning a detailed review if there is little or no available evidence to support it. Early work will investigate whether previous reviews have been carried out on the topic and whether these might be updated.

Searches of Cochrane (www.cochrane.org) and Campbell (www.campbellcollaboration.org) Collaborations, and of the Database of Abstracts of Reviews of Effectiveness (DARE) database, are useful places to look for existing reviews, but these should be considered critically in relation to method and relevance to the topic area of concern. Negotiation with commissioners, and consultation with users, practitioners and other stakeholders, should also inform the focus of the work. An advisory group may be convened to support it.

Reviewers may decide to draft a number of questions around the same topic area, and each sub-question may need to draw on different types of study designs. This is discussed further in section 'Introduction to social care research evidence' above. The scope of the work, and the resources on which it will draw, must all be determined before decisions are made about what type of material will be included in the review (see section 'Determining the types of studies that need to be located (inclusion criteria)' below). SCIE recognises and values five types or sources of knowledge: policy, organisational, practitioner, user and carer, and research knowledge (Pawson *et al.* 2003). Knowledge from these sources may contribute to a number of review questions as well as, or instead of, the effectiveness of interventions, including those concerning the transferability of ideas and processes, economic cost and feasibility. SCIE would always give precedence to knowledge from these sources gathered in the course of structured research, for example through qualitative research or formal case studies of implementation. Although empirical research findings are the bedrock of a review, other knowledge, such as theory and debate, may be important background information to a review, and may be formally included in a narrative section of the review, or collected elsewhere.

As a matter of policy, SCIE includes the systematically collected views and experiences of users and carers within reviews, including reviews that concern the effectiveness of a social care approach or intervention. This is not currently part of standard review practice (for example, of the Cochrane and Campbell Collaborations). Research that includes the views and experiences of users and carers allows an additional and vital perspective on how problems are defined, what helps and what hinders the effectiveness of services, and whether a service is acceptable and accessible, as the most effective of services will fail if people do not use it.

Determining the types of studies that need to be located (inclusion criteria)

When the review questions and scope – topic and materials – have been agreed, they will need to be 'translated' into inclusion and exclusion criteria. Developing these criteria means deciding what types of study concerning which types of topic will be included in the review. Criteria will also need to specify the date and language of included publications. For example, an investigation of the impact of a change in policy is likely to restrict included studies to those published after the policy was introduced. Decisions on what to include or exclude will also determine the volume of work, so a small team may strategically decide to limit included publications to, say, the past ten years, if this is consistent with the aim and topic of the review.

The acronym PICO has been used as a memory aide for drawing up inclusion criteria:

Participants: what types of people or populations are relevant?

Interventions: if the review topic concerns interventions, how are they defined?

Comparisons: is it intended only to use controlled studies? If so, what are acceptable controls (placebo, waiting list, standard approaches)?

Outcomes of interest: what measures of effect is the review interested in? Different studies may select outcomes for ease of measurement, but only studies showing outcomes relevant to the review question, and by implication, valued by stakeholders, should be included (Mullen 2006).

An additional 'S' may also be added, namely:

Study designs and methodological quality criteria may be specified (though they may not be clear until the full text of articles have been accessed).

These criteria are used both to draw up a search strategy (see section 'Carrying out a literature search to locate studies' below) and to determine whether the research papers found during searching are in fact within the remit of the review. Each paper found by searching databases is considered in the light of the inclusion criteria. This is commonly called screening for inclusion (see section 'Screening search results against inclusion criteria'). Criteria are likely to be amended when they are piloted against the material found by searching, because the material arising will cover areas and introduce uncertainties not anticipated in the original question. A review question concerning 'looked after children' will need to incorporate a definition: how long, for example, do children need to be 'looked after' for studies concerning them to be included (assuming they meet the other criteria)? Are children cared for by kin who are not parents included?

Any changes to inclusion criteria must be agreed and communicated by all the reviewers who are taking part in the screening. A table is a useful format for the criteria, and version control needs to be tight, as using the wrong version will necessitate re-screening of possibly large numbers of documents. Once inclusion criteria are agreed, anything which does not meet them can be excluded. It makes sense for the table to be ranked, so that the items on which it is easiest to exclude material (such as date of publication) can be dealt with before more difficult criteria, such as the quality of the control group, are considered.

Carrying out a literature search to locate studies

Searching for literature is now almost entirely carried out online using internet-based databases that have been set up by publishers to store and format entries describing articles about research published in journals. The aim of searching in a systematic review is to find as many potentially relevant items as possible using approaches that are consistent with the time and resources available for searching and screening. Reviews need a coherent, systematic and documented search strategy, as this is a central aspect of the methodology and will affect the findings. No search can expect to identify every piece of literature relevant to a review question, but piloting searches on individual databases, and recording the final search for each database, means that advice can be taken and searches repeated and expanded. Internet access to electronic databases has enabled more of us to attempt searching, but it is a challenging field for newcomers, especially in identifying the best source databases and getting the best out of their varied classification and keyword systems.

The choice of databases to be searched in relation to a particular topic area is an issue (see below for suggestions): it should be possible to test the relevance and utility of a database by running small exploratory searches. Searches usually rely on the use of search terms being present in titles, abstracts and keywords, so retrieval of relevant references is dependent on how the papers have been abstracted and their keywords. If there is a transparent structure to the keywords which can be accessed (such as the Medical Subject Headings (MeSH) system, www.nlm.nih.gov/mesh/), this may help in 'translating' the topic question into keywords used in that particular database. There is usually a trade-off in searching between specificity (very specific searching that may limit the number and range of items retrieved as relevant) and sensitivity (broader searching that may lead to more extensive retrieval which will likely include a higher proportion of irrelevant information).

In general, search strategies are devised by identifying relevant databases and developing strings of terms, linked together with Boolean operators (AND/OR/NEAR/NOT), together with synonyms specific to the databases. A clear review question can be divided into sections for the purpose of developing appropriate search strings. For example, to search for items on the topic of 'day care for children with learning disability', you might use the following:

> (terms for) children OR (other terms for) children AND
> (terms for) learning disability OR (other terms for) learning disability AND
> (terms for) day care OR (other terms for) day care.

It is likely that these terms will need to be varied for each database's taxonomy. Including, as SCIE does, structured accounts of user and carer views of a service or intervention may require additional specific terms – for example, for those describing service use (such as consumer, client, patient).

Searching is a pragmatic activity: if a search identifies 20,000 items and the review team comprises two people working in their free time, it may be necessary to narrow the scope of the review and the inclusion criteria. This could be done by revising the publication date (for example, articles published in the last ten, rather than 20 years), and/or the population and/or study designs of interest. Searching on a particular topic will usually yield diminishing returns over time as numbers of duplicates rise: expert guidance on which databases will give highest relevant references, and when to desist from searching, will be useful to the relative novice (Booth 2008).

It is important to select the most useful databases for the topic in question, and this is an area where topic experts, information specialists and librarians can advise. In the field of social care, the following items are recommended (but the list is not exhaustive):

- Medline or Pubmed (if the topic area touches on healthcare)
- Social Care Online (hosted by SCIE, and referencing key social care and social work journals)
- A general social science database (e.g. IBSS)
- PsycINFO, which has useful and sometimes unique material which crosses social and psychological domains including intellectual disability
- Cochrane and Campbell websites for existing or related reviews. The Campbell Sociological, Psychological, Educational, and Criminological Trials Register (C2-SPECTR) has material relevant to social welfare and social care.
- BL Direct (ZETOC) for journal articles, and BL catalogue for books and grey literature (but note that these British Library resources tend not to have abstracts)
- Topic-specific databases, such as ChildData or AgeInfo. The British Education Index (BEI) and Educational Resources Information Centre (ERIC) may be useful additional databases to questions relating to children and young people. Other relevant

databases may offer ‘grey’ literature: for example CommunityWise, NSPCC Inform, DrugData.

- Independent research bodies may also be good sources of material, e.g. Alcohol Concern, Centre for Research in Ethnic Relations, Joseph Rowntree Foundation, Carers UK, King’s Fund, etc. These can be accessed via web searches.

Checking reference lists from highly relevant articles, hand searching of key journals, and personal communication with experts, and experts by experience, in the field are additional strategies for finding relevant material, including grey literature. These ‘serendipitous’ approaches may be all the more important in finding evidence relating to complex, management and policy issues (Greenhalgh and Peacock 2005).

The search strategy developed to underpin electronic or other searching should be included as an appendix to the eventual review report. It should cover any general limitations applied in the search.

Screening search results against inclusion criteria

Search outputs are usually saved into electronic databases, such as Endnote, Mendeley or Reference Manager, and include title, authors, date and publication journal along with an abstract (if available). The title, abstract and (if access can be secured) the full text are screened against the inclusion criteria to remove those irrelevant to or excluded from the scope of the review. Eligibility should be checked by more than one rater and discrepancies formally discussed. Very often, this stage will throw up queries about what is included and what is not, and the review team may need to agree and record amendments or clarifications to the inclusion criteria.

The studies retrieved, screened and excluded should be recorded and total numbers summarised in a flowchart. This will detail the number of items found in searches; the number of items found by other means (e.g. personal contact, stakeholder input, hand searching, citation tracking); and items excluded at first screening with reasons for their exclusion (for example, date of publication; population; intervention). The items apparently meeting the inclusion criteria should then be retrieved as full texts: these are needed for the later stages of the review, including the analysis.

Full text retrieval is a specialist process, with copyright implications if more than one copy is required. Some full text articles may be accessible online, depending on the host institution’s subscription status, and research reports are increasingly published online. Librarians in academic institutions are a useful source of advice for finding full text items in more obscure journals. SCIE has found the British Library to be an efficient source of retrieval for the remaining items, and in particular for British ‘grey’ literature, which is not available through publisher websites. There, is however, a charge for each item. SCIE has also tended to exclude books, as they are difficult to retrieve, and the key findings from research they may contain are usually available through journal articles.

The flowchart showing the identification and management of research records will continue to record the progress of the review, detailing the number of full-text items retrieved (and those that could not be found); the number excluded on full-text screening, and the number of items which therefore went forward for inclusion in the review.

Occasionally, and particularly where inclusion criteria specify methodological or design features, no studies meeting the inclusion criteria are met. 'Empty' reviews are those which, having specified the types of study they intend to include, fail to find any. The reviewers should comment on the lack of appropriate primary studies and perhaps recommend the focus and design of research which might be commissioned to fill the void. It may also be possible to respond to sub-questions within the review topic. The reviewers should always ensure that their conclusions can be summarised and supported by statements of evidence, so the summary could say: 'There were no good quality experimental studies that addressed the effectiveness of this approach, though there were several qualitative studies that demonstrated popularity with service users and carers'.

Critically appraising the quality of included studies

Once full texts of included items have been retrieved, they will need to be critically appraised for methodological quality. There are many standard measures for assessing the methodological quality of included research studies, but these are both varied and substantially overlap, and what is relevant depends on the methodology used in each type of study. A systematic review should report explicitly how studies were assessed. Reviewers can adapt or design their own appraisal tools, but should include these as review appendices. Aspects of the appraisal of studies included in the review should be recorded as evidence tables (NICE 2009): simple text tables where the design and scope of studies are summarised.

There are several reasons for assessing study quality. These include:

- The review may have set inclusion criteria which demand particular study designs (such as RCTs only) and/or demand that these methods have been followed to a certain degree of competence. A review question concerning the cost-effectiveness of an intervention would normally rely on high quality controlled studies. The study design and methods may not have been clear in the abstract, but the full text can be excluded at this point if the study does not meet quality criteria.
- Aspects of the methodology of studies, such as sample sizes and the clear segregation of intervention and control groups, are a guide to the amount of confidence that can be placed in the study findings. In studies using quantitative outcome measures, aspects such as measures used and the power of the study to detect significant change are critical; robust methodology in mixed method and qualitative studies also remains important. For example, a qualitative study of service user views of an intervention will be compromised if it only accessed respondents from majority ethnic groups or failed to include people who opted out of the service.

- The reader of the review can use the evidence tables to check the details, and assess the credibility and generalisability of findings, of particular studies.
- The evidence tables are useful memoranda for the review report writer. The evidence table is a working paper for summarising the evidence within each study for particular conclusions and recommendations.
- The overall set of results from study appraisal is a useful clarification of the nature and shortcomings of the evidence base and should be reported as part of the review findings.

Study design and execution is closely linked to the aims of the study, and some have suggested that the privileging of well-conducted randomised controlled trials is only relevant to effectiveness and efficacy studies of well-determined and relatively simple interventions. However, the disjunction between trials and other designs may not be that great (Dixon-Woods *et al.* 2005), in that they all aim to reduce bias which may invalidate the accuracy or relevance of findings. Referring to Cochrane (healthcare) Reviews, which depend upon RCTs, Higgins and Green (2006) identify four sources of bias in primary RCTs:

Selection bias: systematic differences in the initial composition of the groups;

Performance bias: systematic differences in the care provided to the two groups, apart from the interventions under investigation;

Attrition bias: systematic differences in dropouts and withdrawals that alter initial group composition;

Detection bias: systematic differences in outcome assessment (e.g. expectancy effects due to un-blinded assessment).

These may all apply in some way to a comparative qualitative study of users' views of social care services: the initial samples may not be representative, the interventions may not be delivered as proposed, people may drop out early so that their dissatisfaction with the intervention is not captured, and the follow-up interviews may be 'led' by the researcher's assumptions.

There is no space here for a full discussion of the comparative merits of quality appraisal tools. Further resources in quality appraisal – checklists, received wisdom, good practice – can be found in chapter 5 of Petticrew and Roberts (2006). SCIE's systematic review guidelines (Rutter *et al.* 2010) contain more detail on issues around quality appraisal. For qualitative research, Spencer *et al.* (2003), Attree and Milton (2006) and Mays and Pope (2000) offer reassuringly complementary frameworks for assessing quality, paying attention to suitability of design, inclusiveness of sampling, clear aims and reporting of method, triangulation and attention to negative (contradictory) cases in theory development, and reflexivity.

Dixon-Woods *et al.* (2007) carried out a comparative study using different approaches and tools to quality appraise qualitative research, comparing ‘unprompted judgement’, a quality framework used by the Cabinet Office, and a Critical Appraisal Skills Programme (CASP) tool. Consistency between different raters – which might tell us that the method of appraisal is sound – was not improved with the use of structured appraisal tools, even though raters were able to relate their judgements (as judgements they clearly are) to explicit factors. The appraisal of quality is not straightforward, and should be carried out independently by two or more people. Where stakes are high, assessors may be blinded to the authorship, institution and journal of the studies, as they may all carry weight and influence assessments. Other items are factual, for example study size and characteristics, but can still be sources of error. Study appraisal is most often hampered by the incomplete reporting of methodology, and this may trigger a low ‘weighting’ (see following section).

Aspects of study methodology and quality for each study included in the review should be recorded in one or more evidence tables. It is appropriate to assess study quality in a separate table for each type of study included and use suitable criteria for each type of study (though these will overlap). One column in the table may attempt to summarise the quality using a transparent coding scheme (of which ‘weight of evidence’ is one example: see below).

‘Weight of evidence’ summary ratings of individual studies will pay due regard to study design and quality appraisal, but they are also likely to be concerned with the content of the material and relevance to the review topic and question, and are therefore explained in the following section.

Critical appraisal of the quality of included studies may be combined with data extraction (see following section below). This is a potential short cut, as the full text papers should be read for both operations.

Extracting data from studies, drawing up evidence tables

Data extraction is a process used to describe and categorise the studies included in a review. It is essentially a process, summarised in keywords of the research team’s choosing, by which studies illustrating similar features can be identified and brought together for analytical purposes. It is the first stage of arriving at a framework for analysis and reporting.

Data extraction:

- familiarises reviewers with the range and content of available material;
- enables the identification and summary of topic areas of evidence;
- facilitates retrieval of studies according to topic areas, methodology or other characteristics, by allowing searching on the data keywords;
- is often a first iteration of a framework for the analysis;

- may identify gaps in evidence (where anticipated topics are not found);
- helps reviewers focus on the content of retrieved studies;
- can be integrated with quality appraisal and weighting of studies.

Teams should devise a tool – a tabular checklist of options within each category, that is review-specific and in which there should ideally be no more than 15–20 categories of importance to the review topic. For example, if the review concerns interventions to promote the well-being of looked after children, sections of the data extraction form may include categories for: type of intervention; characteristics of children in study; outcome measures used. It is most useful to the analysis if the options for each category are comprehensive and the 'Other: specify' option is rarely used. As several reviewers are likely to be involved, the tool may include notes on definitions (which again help in reporting transparently on the review process).

To minimise human error and bias, and to increase transparency, data should be extracted by a minimum of two reviewers who should compare and discuss any discrepancies so that they are resolved. The time needed for data extraction should not be underestimated, as it is a lengthy process.

Programmes such as Eppi-Reviewer (EPPI-Centre Sept 2006) are designed to facilitate online data extraction, as well as retrieval of studies according to their similarities for analysis and synthesis. Eppi-Reviewer will also report on discrepancies between team members, among other functions. However, the basic functions of keywording and searching can equally be carried out by recording the keywords in free fields of the databases of included items held in Endnote or Reference Manager. For example, if the context of the delivery of end of life support is felt to be a useful organising principle, data extraction could identify the context of different research studies as hospital, nursing home, own home and hospice care. In preparing that part of the synthesis that deals with hospice-based care, all the relevant studies can be retrieved as a set.

A cheaper option for the first time reviewer is to attach keywords signifying the specific items so that relevant studies can be identified within a word-processed document. Computer assisted qualitative data analysis (CAQDAS) programmes (such as Atlas TI and NVivo) could also be used for this process.

Reviews will frequently provide a summary rating of the quality of each study, referred to as 'weighting'. This is a shorthand means of synthesising the quality of study design, fitness for its own stated aims and relevance to the review topic. Reviews may refer to, or amalgamate, the evidence by reference to different weights. There are competing methods for attaching authority to research studies, and the protocol for the systematic review should describe how this will be done. An example of a weighting system can be found in Dickson and Gough (2009), and application to health services reviews is discussed in Edwards *et al.* (2000).

The quality appraisal and weighting of studies permits some useful summary statements in the conclusions, known as evidence statements (National Institute for Health and Clinical Excellence 2009). An example might be: 'Two highly-rated controlled studies support the finding that people benefit from intervention X, and no studies were found to show that intervention X had no effect or poorer outcomes than the usual care.'

In a world where there is not enough time for all who need to know it to read the detail, these summary statements can be very useful and accessible, especially to people who are not primarily researchers.

Synthesising the studies and assessing heterogeneity

Data concerning topic content and study quality extracted from the studies will be synthesised in the review. Some of this information is summarised in the evidence tables, which form part of the report. Methods of synthesising studies vary according to the type of data reported. Different approaches apply, so the first step is to disaggregate different types of study, an activity which was probably completed with the drawing up of separate evidence tables during the quality appraisal stage. As many studies used mixed methods, it is quite possible that a single paper may appear in more than one table or synthesis.

This is not the place to give extended guidance on analysis and synthesis. Rather, some observations and further references on meta-analysis of quantitative data and on qualitative data synthesis are offered as an introduction.

Meta-analysis

A quantitative (statistical) meta-analysis may be used to synthesise numerical data. Meta-analysis is 'a review that uses a specific statistical technique for synthesizing the results of several studies into a single quantitative estimate (for example, a summary effect size)' (Petticrew and Roberts 2006). More accurately, it uses several statistical techniques and is not recommended to those without a sound understanding of statistics. There is considerable guidance on statistical meta-analysis (often referred to incorrectly as though it is the only form of 'systematic review') developed for use in other review organisations: for example, see Higgins and Green (2006), NICE (2009) and Littel *et al.* (2008).

The purpose of meta-analysis is to pool the results of studies which address the same research questions using similar outcome measures. Meta-analysis is the statistical process of combining the results of similar randomised controlled trials in order to estimate the likely effect size of the intervention that is being tested across an aggregate of all the samples.

The clear presentation of meta-analyses may hide the actual messiness behind the amalgamation of studies, particularly where interventions are complex. Consider:

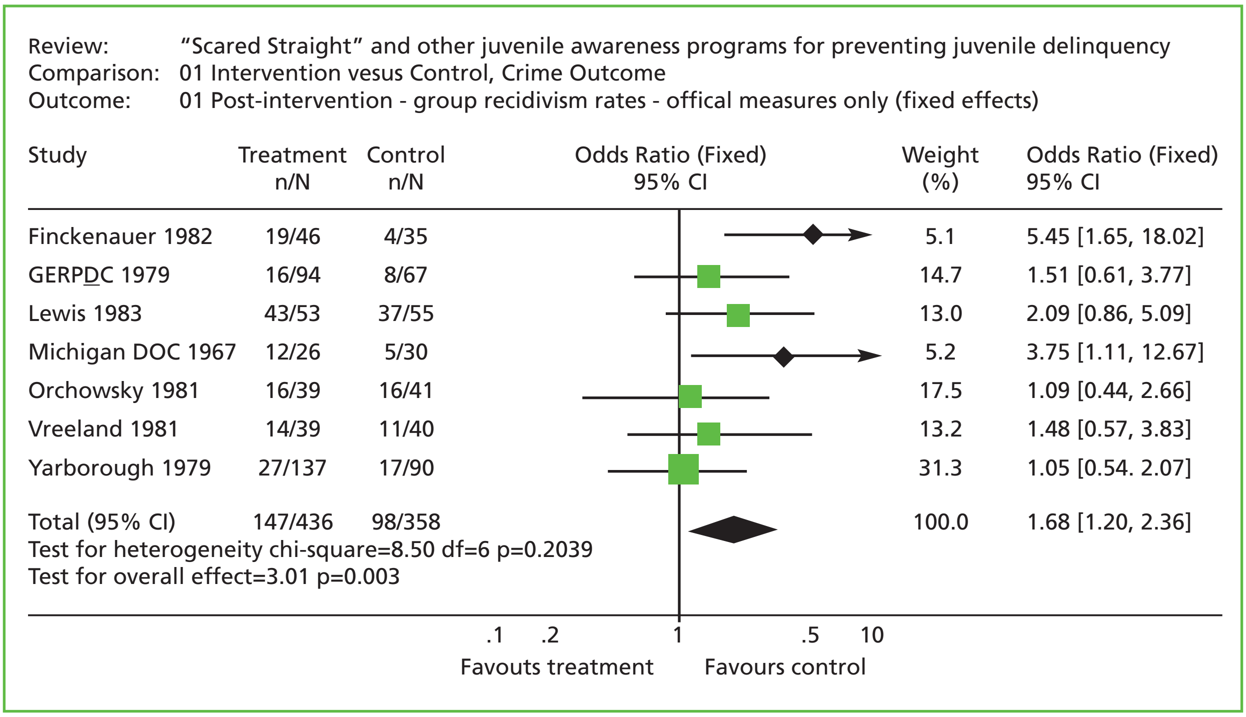
- what appears to be the same or similar intervention is possibly not. For example, a systematic review of Therapeutic Communities (TCs) (Lees *et al.* 1999) combined in a single meta-analysis results from studies of *prison-based* TCs for male offenders undergoing 12-step therapy for substance abuse, with results from TCs attended voluntarily *in the community* by people with personality disorder.

- the 'success' of an intervention depends on what it is compared to and other contextual factors relating to the intervention. For example, NHS Evidence currently rates the evidence for Assertive Outreach Treatment (AOT), a health and social care service, for people with schizophrenia as 'surrounded by uncertainty', although this same evidence has been used to justify significant commissioning within UK mental health services. Studies identified in the Cochrane review (Marshall and Lockwood 1998) used three different controls, of which 'treatment as usual' would be highly varied, especially when it relates to different countries and continents, as in this case. Advantages for AOT recipients were found in relation to housing, employment and user satisfaction with no specific health or symptom-related benefits other than less time spent in hospital. The content of AOT is also disputed and highly variable across studies. These issues are not immediately clear from the presentation of a meta-analysis, and most reviews depend upon narrative explanation to be comprehensive.

Meta-analysis is a process that demands similar studies with similar comparators, and possibly works best where there are clearly defined treatments or processes being tested, and there have been several RCTs using similar measures. The effect under scrutiny might be an improvement on a particular diagnostic scale or pain measure, or on a more proxy measure of health (or cost), such as the time a person is enabled to stay out of hospital (Marshall and Lockwood 1998). It is part of the culture of health research to repeat studies to gain greater confidence in findings. In the complex world of social care, social welfare or education, where RCTs are anyway less common, it is rare to find a number of studies of sufficiently similar design to pool results (Boaz *et al.* 2004). Randomisation is sometimes seen as unethical by social care service providers, so research may be controlled but not randomised: see for example Lewin and Vandermeulen (2009) on re-ablement. Non-randomised studies are thought to be open to bias and will not be included in meta-analyses. However, many topic areas straddle the margins of health and social care, and complex social and psychological research should ideally include mixed methods (Craig *et al.* 2008). It is therefore useful to understand the basic remit and limitations of meta-analysis, represented graphically by the forest plot (or graph).

A meta-analysis shows the range of outcomes from the different trials which meet inclusion criteria; what confidence can be placed in the assertion that the outcomes lie within a certain range of effect; and in summary, the combined average or mean effect size, as though all the people who participated had been put into one large single study sample. A forest plot is appealing because it is easy to read, but it represents the summary of much thoughtful effort, and care must be taken, for example, that studies reported in several papers are not included more than once.

Figure 1: Example forest plot



Review: "Scared Straight" and other juvenile awareness programs for preventing juvenile delinquency
Comparison: 01 Intervention vesus Control, Crime Outcome
Outcome: 01 Post-intervention - group recidivism rates - offical measures only (fixed effects)

Study	Treatment n/N	Control n/N	Weight (%)	Odds Ratio (Fixed) 95% CI
Finckenauer 1982	19/46	4/35	5.1	5.45 [1.65, 18.02]
GERPDC 1979	16/94	8/67	14.7	1.51 [0.61, 3.77]
Lewis 1983	43/53	37/55	13.0	2.09 [0.86, 5.09]
Michigan DOC 1967	12/26	5/30	5.2	3.75 [1.11, 12.67]
Orchowsky 1981	16/39	16/41	17.5	1.09 [0.44, 2.66]
Vreeland 1981	14/39	11/40	13.2	1.48 [0.57, 3.83]
Yarborough 1979	27/137	17/90	31.3	1.05 [0.54. 2.07]
Total (95% CI)	147/436	98/358	100.0	1.68 [1.20, 2.36]

Test for heterogeneity chi-square=8.50 df=6 p=0.2039
Test for overall effect=3.01 p=0.003

Source: Petrosino *et al.* 2003

Features of the forest plot are illustrated by the Petrosino *et al.* (2003) Campbell review update (of an earlier Cochrane Review) of school programmes in which children felt to be at risk of offending were taken on tours of adult prison facilities (see Figure 1).

In this example, the expected deterrent effect was not demonstrated, and in relation to the key outcome of future criminal behaviour, those in the control group fared better, suggesting the programme was actively harmful. Each of the seven included studies (referenced on the left of the diagram) is represented as rows in the forest plot with mean effect sizes plotted for each study across a vertical axis by a rectangle. The position of each shape on either side of the central line, the 'line of no effect', shows whether the study results favoured the treatment or the control, and the combined studies are represented by combining the samples in an additional line. It is important to note that each study's mean results are supplemented by a horizontal extension of the shape to show the confidence interval, in recognition that the outcomes of a study are a likely, but not precise, indication of the relationship of two variables (such as a treatment and improvement in health) as it might apply to any population. Where the horizontal confidence interval crosses the line of no effect, it is possible that the intervention had no discernible effect when compared to the control. Relative sample sizes are roughly represented by the relative sizes of the shaded rectangles. The combined studies, as though all the samples and effects were added together, are shown at the bottom by a diamond.

In social care and social work, there is frequently insufficient data of the required type and quality to form the basis of a meta-analysis. Where controlled studies exist, samples may be small. Petticrew (2003) concludes that a quantitative meta-analysis of many small studies is more likely to detect a false positive effect (by combining sample sizes and so increasing power to detect), while narrative analysis, bringing together and acknowledging the disparity of its sources, is more likely to falsely conclude that there is no effect.

Boaz *et al.* (2004) report on 28 reviews undertaken for government: only two within the set could use the technique of meta-analysis. As discussed in section 'Introduction to social care research evidence' of this paper, there are anyway other research designs that are extremely important to social care and social work review questions, and the precedence accorded to RCTs and to meta-analyses is possibly misplaced in that context. However, the synthesis of all types of quantitative study and data is beyond the scope of this paper: further remarks on synthesis concern only qualitative studies, as the inclusion of these in reviews has been controversial.

Qualitative data synthesis

Narrative synthesis, giving at least a description of the studies included and the overall findings, should be a feature of all reviews even where the main synthesis focuses on controlled quantitative studies (Noyes *et al.* 2008). This is useful advice: one of the difficulties with meta-analysis is that there is often little detail on the interventions assessed. This guidance from the influential Cochrane Collaboration also suggests a key role for qualitative studies to enhance reviews of effectiveness by offering an understanding of the 'experience of those providing and receiving interventions … and factors that shape the implementation of interventions' (Noyes *et al.* 2008).

Qualitative data synthesis may have two broad aims: (i) *integrative*, where the purpose of the analysis is to combine the findings of different studies; and (ii) *interpretive*, where the purpose is to develop further or 'higher-order' explanatory concepts or theories consistent with the different studies or data sources (Dixon-Woods *et al.* 2005). These broad approaches are not confined to the synthesis of qualitative data, as the authors demonstrate: they characterise the synthesis of many types of study, but are perhaps more explicit in qualitative analysis. Qualitative data synthesis (QDS, also called *qualitative evidence synthesis* by Noyes *et al.* (2008)) is not, then, just a literature review, but should generate 'a greater degree of insight and conceptual development that constitutes a fresh contribution to the literature and … an understanding of why (people) feel and behave the way they do' (Noyes *et al.* 2008).

Qualitative studies may address several different aims or questions within a review, so a first step may be to use the data extraction tools to aggregate material relevant to different topics. (Popay *et al.* 2006) have identified four main elements to narrative synthesis:

- developing a theory of how the intervention works, why and for whom – the aim of which is to inform decisions about the review question, inclusion criteria and interpretation of study findings. A realist approach (Pawson *et al.* 2005), (see section below 'Process, implementation and the realist approach: pushing systematic reviews further') may be useful here.
- developing a preliminary synthesis of findings of included studies – the aim of which is to organise findings in order to be able to describe patterns across included studies, including the direction of effects;
- exploring relationships in the data – the aim of which is to consider factors that might explain differences across study findings;
- assessing the robustness of the synthesis – the aim of which is to assess the strength of the evidence included in the review for drawing generalisable conclusions.

In practice, reviewers will move in an iterative manner among the activities making up these four elements.

QDS has been given substantial impetus by the work of Sandelowski (Sandelowski and Barroso 2007) and Dixon-Woods (Dixon-Woods *et al.* 2006), both of who include worked examples. Researchers considering QDS should consult these sources and consider whether to employ some of the techniques. For example, Dixon-Woods *et al.* (2006) sample within the total number of retrieved studies (rather than reading and coding all of them) and their quality criteria exclude studies only if they are 'fatally flawed'. SCIE has also developed a worked example of systematic synthesis (Fisher *et al.* 2006) on the synthesis of research data on older people's views of hospital discharge. SCIE's worked example builds on the staged approach of Noblitt and Hare (1988):

1. The findings from the primary studies, such as the meanings reported to researchers (sometimes called *first-order* interpretations). For example, older people may say they perceive doctors and nurses as having more expertise in health and illness.
2. The constructs and interpretations that primary researchers place on these findings (*second-order*). For example, that this creates dependency on staff for information (a researcher construct).
3. Explanations and hypotheses developed by reviewers arising from second-order interpretations (*third-order*). Trust is undermined when people perceived as experts do not agree and that anxiety increases when access to medical expertise is reduced (a construct arising from synthesis).

The process of working through these three stages is as follows:

a. The reviewers use the material provided by data extraction forms to identify findings and concepts. This process resembles a method of analysis known as grounded theory, in that it involves identifying conceptual categories and the studies (or extracts from studies) that support them.

b. Core findings and concepts are compared across studies (sometimes this process is called ‘translation’ or ‘reciprocal translational analysis’); a grounded theory approach is again relevant, in that the process resembles that of seeking similarities and differences between findings and concept. The process can also involve noting where expected similarities are not found and trying to explain why (sometimes called ‘deviant case analysis’ in grounded theory, akin to ‘refutational analysis’ in QDS).

c. In this way, initial broad coding categories (e.g. participation of older people) are identified and tested until it is clear they are central.

d. The reviewers should maintain an audit trail, linking synthesis statements to supporting studies or extracts and should cite the supporting studies or extracts in the account. Again, software for computer-aided qualitative data analysis can assist with this.

e. The synthesis and the line of argument that links findings and concepts should then be written up in such a way as to make the process of analysis as transparent as possible; in the worked example provided by the SCIE example, the synthesis is tabulated in three columns showing the first-, second- and third order stages.

None of the processes described above is a blueprint for qualitative data synthesis. QDS should take account of the processes described here and should demonstrate a transparent approach that permits the reader to interrogate the processes and potentially to replicate them. It is very helpful if QDS is undertaken by at least two reviewers, working independently at key stages, for example to identify core findings and concepts.

Reporting and disseminating

The transparency of approaches to data analysis and synthesis requires that review methods are substantially reported and may comprise most of the review. Making the protocol available separately online may help to ‘unclutter’ the report. Using plain language and clearly signposted conclusions are important if the audience consists of policymakers or other non-researchers; but there is clearly a trade-off between plain speaking and the loss of subtleties, such as the reporting of the strength of evidence of key studies. Evidence statements (see section ‘Extracting data from studies, drawing up evidence tables’ above) attempt a compromise.

Limitations of the review, including publication biases, and limitations of included studies and analysis should be reported. One task of the review is to report on assessment of the evidence base for this topic. To be of maximum utility, the review could consider applicability, utility and generalisability of the findings to policy and practice, as well as economic implications of findings and possible recommendations. However, it may be inappropriate for researchers to elaborate too far on practice issues and recommendations, particularly if the review is only one source of evidence within a project or programme: a project advisory group of varied stakeholders would have a better chance of formulating recommendations that were likely to be relevant to practice contexts.

It is increasingly frequent, with the Cochrane and Campbell promotion of review methodology, to find that systematic reviews cannot demonstrate an effect. Absence of a way of assessing outcomes does not necessarily mean there is no effect (Petticrew 2003). If the topic of the research is important, more primary research may be needed, and this will form part of the recommendations.

Dissemination of findings is a vexed topic, often referred to as ‘knowledge transfer’. While there is no scope here for further remarks on dissemination, it is clear that most reviews do not report in a format – or in a context – which is easily accessible to practice, policy and even research audiences. This in itself is a challenge for systematic reviewers.

METHODOLOGICAL ISSUES AND CHALLENGES TO THE METHODOLOGY OF SYSTEMATIC REVIEWS

There is a range of methodological issues that arise in the conduct of systematic reviews, particularly when the methods are applied in the context of social care and social work. A few key ones are selected here for comment.

The first two concern the process of reviewing: the involvement of stakeholders in the review process and considerations around topic selection. The third issue concerns the status and shortcomings of the use of randomised controlled trials in systematic reviews. This is followed by two sections considering the apparent schism between quantitative and qualitative research designs and the use of realist synthesis and other frameworks to make better use of research relevant to implementation. The final two sections explore some caveats concerning the interpretation and use of systematic reviews, especially by policymakers.

Involving users, carers and other stakeholders

There are two broad ways in which service users and carers can be involved in systematic reviews. First, research that explicitly concerns their experiences and views can be specifically included in the review. Second, users and carers can be involved in, or lead, the process of reviewing.

In general, the policy imperative encourages public involvement. Research that ignores minority groups and/or those with complex characteristics is potentially damaging, since findings reporting a particular range of impact on the majority are very likely to be extrapolated to these groups. Data on minority populations can be explicitly targeted by including appropriate terms in search strategies, and the data extraction process can be used to disaggregate findings specifically arising from this population. Even if there are few studies to draw on, the reviewers will be able to point out this omission in the evidence base.

People engaged in some way with the research question, whether as practitioners or service users or carers, should be included in the review process. Users and carers can bring a fresh perspective to framing the review question, the design of inclusion criteria, outcome measures and quality assurance, as well as guiding the analysis framework (Fisher 2005). SCIE maintains and updates a collection of examples of service user and carer participation in systematic reviews (Carr and Coren 2007). Service user involvement in determining the review questions should increase the topic relevance: ‘Objectivity is not a prerequisite for valid evidence (and can even be harmful in some circumstances)’ (Glasby and Beresford 2006).

Where possible, commissioners of reviews should be involved in their conduct, particularly in decisions about scoping and inclusion. It is helpful for all if commissioners are kept aware of difficulties encountered and strategic decisions taken during the production of a review.

Cochrane and Campbell reviewers are largely unpaid, so the topic selection is likely to be determined by their interests. There is no current mechanism by which this increasingly important aspect of knowledge production could be more democratically focused and determined: funders may or may not consult with other stakeholders to decide on topics for review.

Topic selection and proportionality of effort

The selection of a topic for a systematic review will itself require some knowledge of the field: while users, carers and practitioners bring practical knowledge of the context, preliminary exploration of the evidence base is also desirable. Pre-scoping searches, and/or the systematic mapping techniques utilised at Eppi-Centre (Gough *et al.* 2003) and SCIE (Clapton *et al.* 2009), are a means of identifying where evidence exists, which warrants the investment of time and effort required for a systematic review. Reviews that exclude all but randomised controlled trials may find their interesting topic melt away if there are few trials or they are poorly conducted (see section ‘Screening search results against inclusion criteria’ on empty reviews). It is very useful to have an idea of the extent of data sources in order to plan and resource the work – or to abandon or revise the topic where there is too little or too much material.

It is unusual for the end users of knowledge (for example users, service providers and commissioners, or social work educators) to be involved in framing review questions, and there are many issues of ‘ownership of knowledge production’ (Braye and Preston-Shoot 2007) which discourage the use of reviews by the people they are meant to influence. In general, it appears that publishers of academic journals are less interested in discovering and publicising what harms or what performs indifferently than they are with ‘what works’, a phenomenon referred to as ‘publication bias’. However, there is increasing recognition that some widely available interventions may not just be ineffective, but harmful. Pharmaceutical companies are sometimes accused of failing to release details of studies which show negative or indifferent results for their products. The formal quality appraisal of studies included in reviews aims to discover and investigate why studies included in a review show conflicting outcomes: reviewers may well ask themselves how they would have carried out each study.

(Zwi *et al.* 2007) conducted a review of widely available school-based education programmes for the prevention of child sexual abuse. They found 15 trials, several of which reported harmful outcomes, including measurable increase in levels of anxiety in the children exposed to the programme (Zwi *et al.* 2007). The primary studies also focused on outcome measures such as knowledge: the retention of knowledge (though no study involved more than 12 months follow-up); and behaviour change (itself measured by simulated abduction situations). Incidence of sexual abuse, or the reporting of sexual abuse, could not be used as a reliable outcome measure. The realist approach (see section ‘Process, implementation and the realist approach: pushing systematic reviews further’ below) might have questioned whether anxiety was part of the mechanism of change and at what level it should be regarded as dysfunctional. Discovering harm, or lack of effect, should be seen as a legitimate and useful outcome of reviews.

Sources cited by Petticrew and Roberts (2006) suggest that reviews can take a median of 1,110 hours to complete with a wide range either side of this. This is a big investment of time, and delivery dates can confound the intention to use the results to achieve policy imperatives. Reviewers increasingly try to satisfy the policymakers' call for more rapid reviews by taking short-cuts in the review process, for example by reducing the range of searches or tightening inclusion criteria (Abrami *et al.* 2010). The Eppi-Centre suggests that the cost of a systematic review ranges from £50,000 to £80,000 (Oakley 2003), though the scope of the question and availability of evidence is very variable, and so it is difficult to generalise about cost or time. A review may not be worthwhile if few quality studies exist. A first step is always to seek existing reviews of the topic, since updating an existing review is clearly cheaper.

The status of randomised controlled trials in systematic reviews in social care

We have argued that systematic reviews can include any type or design of research study, yet in many contexts, the use of the term is understood to mean an approach to synthesising randomised controlled trials (Guyatt *et al.* 1995; Davies and Boruch 2001). Good RCTs are considered the least biased means of exploring the effectiveness or efficacy of an intervention, and, while they cannot answer every question we might wish to review, their more extensive use in social care and social work studies is surely to be encouraged.

At its most essential level, an RCT takes a sample of people from a single pool of participants and allocates them by chance either to a group who receive an intervention, or to a group receiving a control (often a placebo "dummy intervention"), or treatment as usual. Where possible, both the participant and the researchers assessing participants are 'blinded' to which arm the person being assessed is in. Blinding is a mark of quality in RCTs, as it eliminates bias borne of assumptions, expectations and preferences; but it is of course far easier to maintain blinding if the intervention is a pill, and almost impossible if the intervention is a complex psycho-social intervention, such as CBT groupwork, which is 'co-produced' by the service user.

Randomisation is comparatively rare in social care, and this is a challenge for reviews which aim to focus on whether an intervention is effective. Reasons why this may be so are discussed briefly in the section 'Introduction to social care research evidence' above. RCTs are expensive to implement, and social care research is less well-funded than health research (Marsh and Fisher 2005). In addition to the resourcing of trials, the pragmatic issues of design and implementation in social care contexts have not had the attention they deserve. Cluster randomised trials – where whole care homes or schools host an intervention and other care homes are the controls – could be more frequently used to set up trials in social care. Such designs require sophisticated statistical techniques, and there may be a shortage of statisticians in academic social science departments. But pragmatic and resource constraints may not account for the range of criticism of trials in social science epistemology.

Epistemological commentary on RCTs from social science includes criticisms that:

- They focus largely on effectiveness, measured through differences in identified outcomes (in the intervention and control groups). Service user and practitioner experience is often not reported in such studies.
- They reveal little about why and for whom an intervention is effective, or what contextual factors play a role.
- They are often felt to be inappropriate for the evaluation of complex social interventions, as they do not encompass a theory of the mechanisms of change, being preoccupied with the inputs (the intervention, which is often not well described) and the outcomes.
- RCTs do not pay sufficient attention to the context in which interventions operate, which in the UK may reflect their use within a National Health Service held to be relatively undifferentiated. The social care sector recognises its own diversity and questions the generalisability of RCT results, which do not discuss contextual issues.

The shortage of RCTs in social care, social work and social welfare literature has both inhibited the development of systematic reviews and encouraged social scientists to work with a wider definition of evidence and develop methods for synthesising other types of data. One of the flaws inherent in the experimental process captured by RCTs is that the timeframe of research studies is relatively short: yet many of the interventions important in health and social care, such as therapeutic interventions in foster care, have long-term impacts, where a well-conducted cohort study might be rather more appropriate to the research question (Slavin 1995). It is instructive to recall that some of the greatest insights in disease prevention, such as the link between smoking and the development of cancer (Doll *et al.* 2008) came from long-term cohort studies. Petticrew and Roberts (2006) argue for the explicit inclusion of types of study which appear suited to the research question, suggesting ‘typologies’ as a less loaded categorisation than ‘hierarchies’.

From a social science perspective, (Pawson 2004), questions whether the RCT methodology in itself is ever a marker of quality in its own right: RCTs can be as poorly conducted and reported as any other research design and may be excluded from reviews on this basis. Both systematic review and RCT reports frequently lack adequate description of the intervention, and, focusing on outcomes, cannot break down the active components of an intervention and what effect each may have. Pawson’s Realist Synthesis (Pawson 2002) tries to break down the implied process between intervention and effect into stages which represent the implied theory of change. This is an appealing approach, but it cannot really be seen as systematic without clearer details of how material to evidence these processes is identified (Pawson *et al.* 2003). His critique however stands: RCT methodology need not, but tends to, omit the significant theoretical steps by which an intervention achieves its outcome. The resulting outcome may be positive or negative, but we will not know which part of the programme failed to work as anticipated, nor why.

Other potential flaws of RCTs also concern the conduct of the study in the practice context. (Boaz and Pawson 2005), examining a review on the mentoring of young people felt to be at risk of offending, found that some of the RCTs investigating mentoring followed up only those who had completed a set number of sessions (rather than all admitted to the trial, known as an 'intention to treat' analysis). Since waiting lists were a common problem, this excluded from the results those who grew tired of waiting, or who disliked the intervention and failed to continue with it. Lack of patience and ADHD are probably typical of the intended participants, so the analysis quite possibly screened out the intended beneficiaries and the positive results of mentoring applied to young people who were not the ones for whom it was designed. Standards on the reporting of trials have been developed as a means of disclosing these flaws (Moher *et al.* 2001), but the quality of RCTs in their own terms certainly cannot be assumed, and the review's evidence tables, which give details of included studies, should make flaws explicit.

Paradigm wars? The case for including qualitative and quantitative research in reviews

In the social sciences, there has been polarisation between those championing quantitative and qualitative research, and this may be a developmental phase (Gough and Elbourne 2002). The methods of quantitative research, whereby outcomes must be measurable, and the randomised controlled trial is seen as the experimental method least likely to deliver biased results, are clearly most suited to studies concerning the effectiveness of interventions that can be delivered in a standardised fashion. Many social care interventions are not standardised, and in cash-strapped services, there are rarely choices available which could be experimentally trialled.

However, social scientists have tended to criticise the ideology, rather than the logistics, of RCTs as a basis for addressing systematic review questions (Gough and Elbourne 2002): the sense in which the methodology constrains the questions addressed by research, rather than the practical feasibility of RCTs. The status of RCTs and systematic reviews in the social sciences is lower than it is in healthcare, although reviews are increasingly combining RCTs and quantitative and qualitative research and refining methods to do so (Pearson 2004; Dixon-Woods *et al.* 2005; Harden and Thomas 2005). Intervention studies do have a tendency to concentrate on the input (the intervention) and the outcomes (results), ignoring access and implementation issues, context, acceptability, user satisfaction, sustainability, etc. Conversely, purely qualitative approaches, which record views and perceptions, may well be missing an opportunity to add more objective measures and designs.

Another way of characterising this apparent division between types of systematic review and reviewers is to use the terms “formative” and “summative” in relation to evaluation. A “summative” evaluation is a method of judging the worth of a programme at the end

Box 2: Formative and summative evaluations

Formative evaluation includes several evaluation types:

- **Needs assessment** determines who needs the program, how great the need is, and what might work to meet the need
- **Structured conceptualisation** helps stakeholders define the programme or technology, the target population, and the possible outcomes
- **Implementation evaluation** monitors the fidelity of the programme or technology delivery
- **Process evaluation** investigates the process of delivering the programme or technology, including alternative delivery procedures

Summative evaluation can also be subdivided:

- **Outcome evaluations** investigate whether the programme or technology caused demonstrable effects on specifically defined target outcomes
- **Impact evaluation** is broader and assesses the overall or net effects – intended or unintended – of the programme or technology as a whole
- **Cost-effectiveness and cost-benefit analysis** address questions of efficiency by standardising outcomes in terms of their monetary costs and values
- **Secondary analysis** re-examines existing data to address new questions or use methods not previously employed
- **Meta-analysis** integrates the outcome estimates from multiple studies to arrive at an overall or summary judgement on an evaluation question

Source: Social Research Association website

of the programme activities. The focus is on the *outcome*. A "formative" evaluation is a method of judging the worth of a programme while the programme activities are forming, or happening, and focuses on the *process*. Both types of approach may be appropriate to different types of research question, as Box 2 illustrates.

Process, implementation and the realist approach: pushing systematic reviews further

Pawson (2001) and Pawson *et al.* (2005) seek to break down the conceptual vacuum, known elsewhere as the black box (Pope and Mays 1993), between input and output by developing a theory of change for complex social interventions. By breaking down the constituent stages of what is supposed to happen, one can then seek evidence for each stage and consider deviations from the intended course of events. This has been written up by Pawson and colleagues as realist review or realist synthesis. The type of evidence available to support each stage need not be research-based, let alone quality appraised (Pawson *et al.* 2003). A worked example of realist synthesis in relation to Megan's Law (by which communities would be notified of the presence of sex offenders) in the USA is available (Pawson 2002), but is not considered in more detail here because the approach, while interesting and useful, does not meet the standards of systematic review: it does not use 'transparent procedures to find, evaluate and synthesize the results of relevant research' (www.campbellcollaboration.org). It is, for example, not always clear how the research studies for the realist synthesis are identified. This is no small point when the aim of systematic searching is to do one's best to find all the past research which might help to answer the question.

However, a subsequent paper (Greenhalgh *et al.* 2007) builds on a systematic review by extracting narrative data from the studies included in their Cochrane review on the efficacy of school feeding programmes. A realist review framework was used to consider the process and mechanisms by which each programme appeared to be, or not to be, effective, relying on data reported in, and extracted from, the original studies. For example, evidence was found from many trials that feeding programmes were most effective where the target group had been identified as having clear nutritional deficiency (usually insufficient calories), and where measures were taken to ensure that the food was consumed. Study designs that relied on the exclusion of some children from feeding, so that they acted as a control group, were not implemented by school staff, who found them unworkable or unethical; or by children, who shared the food with friends. Leiberman's study replaced individual randomisation by cluster randomisation, running the feeding programme in some schools but not others to show effectiveness (Lieberman *et al.* 1976). The realist perspective allowed the outcomes to be reviewed against process and contextual factors: a 'realist' perspective being applied to the material gleaned from studies, which were all predominantly trials, in order to ask, 'what appears to have contributed to the outcomes reported in this trial?'

A realist review used in this way is a useful approach to the analysis of materials identified through systematic processes and suggests that traditional systematic reviews may not be making the best use of their studies: but it is doubtful whether this exposition of a realist review is entirely differentiated from the interpretive school of narrative analysis. Both apply a theory-driven evaluation approach using the data to extrapolate and test a theory of implementation, process and impact. Both contribute to a general shift towards the use of process evaluation alongside outcome evaluation to 'clarify causal mechanisms and identify contextual factors associated with variation in outcomes' (Craig *et al.* 2008). Most social science evaluations are of complex interventions that benefit from this dual approach, and most readers of reviews concerning social care and social work are fundamentally interested in implementation issues. Roen *et al.* (2006) carried out further analysis on existing systematic reviews of the effectiveness of interventions aiming to reduce unintentional injuries in children and young people, with a view to identifying data on implementation, from within the studies. Valuable data about the context in which such initiatives are implemented and the type of factors that might impinge on implementation were discovered. Implementation data is often buried within effectiveness studies: Roen et al (2006) recommend that researchers, commissioners and journal editors with an interest in evidence-based public health should be encouraged to consider implementation issues in the design of intervention studies.

Reviews can be misleading and uncertain

Reviews are the best available method to answer questions about the effectiveness and acceptability of interventions, but they do not always deliver answers. The sheer complexity of interventions in the field of social care may militate against being able to capture all effects (Petticrew 2003). A good example of complex interventions are the Sure Start programmes, which have been subject to a great deal of evaluation. A recent search conducted within SCIE found 26 evaluation reports from the national evaluation of Sure Start. It is unlikely that a systematic review process of such complex interventions, implemented within different timeframes, designed to meet disparate local need, delivered by different staff and targeted with different levels of success at varying local populations, would deliver a reliable verdict on the programme. ‘Few relevant outcome evaluations – randomised, controlled or otherwise – of major UK social programmes have been carried out’ (Pettigrew 2003), and policymakers have only in recent years begun to focus on ‘outcomes’, some of which are likely to be difficult to evidence and measure. Some of the outcomes selected to demonstrate improvements in adult social care (for example improved health and emotional well-being, improved quality of life, making a positive contribution, choice and control, freedom from discrimination, economic well-being, personal dignity) (Department of Health 2006) may only be demonstrable by proxy measures.

Where complex interventions are evaluated by rigorous methods such as controlled studies, the lack of contextual, observational and process data is an impediment to practitioners considering their implementation. Unless there are very many studies showing positive outcomes (itself unlikely), the generalisability of the findings – whether the same results might be achieved in any setting – is in doubt. This is a concern, which is felt more keenly within the social sciences and social care research than it is within health research.

Reviewers determine the scope of the review and may choose to categorise a range of interventions as similar (although this is clearly a matter of judgment), particularly where studies from very different international contexts are included. Smedlund *et al.* (2007) reviewed six trials of CBT for men using violence against women. The differences between the interventions, and major differences in the control treatment to which each was compared, did not permit any useful conclusion, except that more primary trials were required.

Similar controversy arose in a review of psychotherapeutic interventions for use with children: Craven and Lee (2006) brought together a number of psychodynamic therapies for emotionally disturbed looked after children which were, it later transpired, too heterogeneous to have been combined. A critique published subsequently alleged that one of these therapies had been shown to be harmful to children (Pignotti and Mercer 2007); and, arguably, they were all too different to be assessed as a set. Small sample sizes in this topic field may encourage the practice of combining therapies of this nature in a single review.

Finally, reviews can be misleading when the 'user' of reviews does not have the capacity or time to assess the subtleties of evidence. It would be difficult to fully explore or evaluate the conclusions of a review without the narrative tables explaining the characteristics of the studies used, which is why these are an important component of systematic reviews. However, decision-makers may not have time to investigate these.

Using reviews as a basis for policy

Systematic reviews are fraught with pitfalls as a basis for policy. Boaz *et al.* (2004) identified tensions between methodological quality and making the best use of the available research, a common problem in social science and social care. The scope (breadth and depth) and inclusion criteria of the review can be set at different points and may be re-negotiated when ambiguous material surfaces, or may continue by default, despite demonstrable inadequacy, because no-one challenges them. While transparency of method and reporting are the defining features of a systematic review, it may be difficult to convey methodological and analytical subtleties to the civil servants who can commission new services.

Subtle conclusions derived from the esoteric (though not so objective) science of reviewing are not easily translated for users of reviews. Reviews can also represent a means of justifying policy to the general public. It has been argued that the lack of understanding of research and aggregations of research by the public excludes them from important public debates such as prevention and management of Bovine Spongiform Encephalopathy (BSE) (Gough and Elbourne 2002).

The reporting of science in general, and reviews in particular, is often flawed. Petticrew (2003) remarks that reviews which go to extreme lengths to conclude that 'good evidence is currently lacking' will not be warmly received by civil servants, especially if large amounts of money and time have been spent. Researchers may feel pressured to shape their findings into practical recommendations. In the study of five systematic reviews of mentoring for troubled young people (referred to above (Boaz and Pawson 2005)), the link between review findings and recommendations to policymakers was examined and found wanting. Quality of evidence, and the content of change shown by meta-analysis, was variously reported either as supporting widespread use of mentoring, or as justifying its rejection as an approach to changing behaviour in troubled youth. Without cost-effectiveness analysis, positive change 'caused' by the intervention might anyway be too marginal to the young people's future development to justify the cost. Some evidence was found that abused young people could suffer actual harm from mentoring. Overall, while researchers no doubt want to be helpful and to translate knowledge into simple messages and practical recommendations, 'fuzzy inferences are dressed and delivered as hard evidence' (Boaz and Pawson 2005 188). Perhaps there is some comfort then in the finding (Lavis *et al.* 2005) that decision-makers hardly ever quote reviews as a source of guidance.

APPLICATION TO SOCIAL SCIENCE, SOCIAL CARE AND SOCIAL WORK

Some of the difficulties of exploiting the evidence base for social care and social work using review techniques are becoming less relevant as methods of appraisal and synthesis are developed for the inclusion of qualitative and narrative material. The comprehensive systematic review guidance from CRD at York (NHS Centre for Reviews and Dissemination 2001) identifies methods for dealing with such material. The Eppi-Centre (at the Social Science Research Unit at London University's Institute of Education) has also been a key player in applying methods of qualitative synthesis to studies of the views of users (Harden *et al.* 2004). SCIE has produced a systematic review of older people's views on hospital discharge (Fisher *et al.* 2006), and has a declared interest in its own commissioning of reviews in drawing on knowledge from organisations, practitioners, service users and carers and the policy community, as well as from researchers. Research reviews and studies that have purposively sought out other data sources and modes of enquiry to complement and reinforce the findings of experimental methods are becoming increasingly common.

Policymakers too should welcome the inclusion of evidence concerning implementation factors and stakeholder views. The funding of an effective intervention that had limited generalisability because it required a particular infrastructural context, or of an intervention that was highly stigmatised among the group it was intended to benefit (such as a treatment programme for crack users delivered in services designed for opiate users (Weaver *et al.* 2007)) would be a great waste of public money. An intervention proven effective in trials but disliked by potential users will likewise prove worthless if people avoid it.

The following table includes some examples of the use of systematic review methods in social care, social work and social work education. These have been selected (relatively unsystematically) to illustrate something of the range of topics which have been addressed by systematic reviews relevant to social work, social work education and social care. It is notable that many found poor evidence, some could not reach a conclusion on the utility of interventions, and some merely recommend strategic research to address the gaps found. It is also relevant to note that activities found to be unhelpful or even harmful may be continued for other reasons (such as popular support, or lack of alternatives coupled with a political need to be seen to 'do something').

Many of the study topics cross disciplinary and professional areas of interest, especially those concerning psychosocial interventions which may be delivered by health or social work personnel. This would be reflected in the choice of databases, journals and topic experts included in the search strategy.

The table omits the very many useful applications of systematic review methods to healthcare, public health, community safety, education and criminology.

Table 1: Selected systematic reviews in social care, social work and social work education

Study reference	Selected summary findings
Attree P (2004) Growing up in disadvantage: a systematic review of the qualitative evidence, Child: Care, Health and Development, 30, 6, 679–689.	Children and young people describe aspects of family relationships, friendships and neighbourhoods that help to mitigate the impact of disadvantage on their well-being. However, their accounts demonstrate that such resources are not universally experienced as supportive and protective: poverty, by its very nature, compromises the worth of these resources.
Gough D, Oliver S, Brunton G, Selai C, Schaumberg H (2001) The effect of travel modes on children's mental health, cognitive and social development; a systematic review. Report for DETR. , EPPI-Centre, Social Science Research Unit, London.	Studies identified found to be poor in methodology and topic coverage. Strategic research priorities recommended, including effect of modes of travel and components of the travel experience (physical activity, nutrition and diet, social, cognitive and environmental experiences) on outcomes (including readiness to learn, lethargy, exhaustion, academic performance, social and cognitive skills); more diverse research designs and wider involvement of stakeholders in research, including young people.
Petrosino A, Turpin-Petrosino C, Buehler J (2003) 'Scared Straight' and other juvenile awareness programmes for preventing juvenile delinquency (updated review), Campbell Collaboration Systematic Reviews, 2004.2.	Programmes like 'Scared Straight' are likely to have a harmful effect and increase delinquency relative to doing nothing at all to the same youths. Given these results, this programme cannot be recommended as a crime prevention strategy.
Macgowan M (2004) Psychosocial treatment of youth suicide: a systematic review of the research, Research on Social Work Practice, 14, 3, 147–162.	The number of interventions for use with adolescents who have attempted suicide or have suicidal ideation has increased in the past ten years, but the research evidence remains weak. Recommendations are offered to strengthen the research, and the role of social workers in using these findings and advancing the research is discussed.

Table 1: (continued)

Study reference	Selected summary findings
Barlow J (1999) Systematic review of the effectiveness of parent training programmes in improving behaviour problems in children aged 3–10 years, Health Services Research Unit, University of Oxford, Oxford.	High quality studies on the effectiveness of parent-training programmes in improving the behaviour of children between the ages of three and ten years are scarce. Those available showed that: • Group-based parent-training programmes have a positive impact on the behaviour of children between the ages of three and ten years. • Group-based parent-training programmes are more successful in the long term in improving the behaviour of children compared with methods that involve working with parents on an individual basis.
Barlow J, Coren E, Stewart-Brown S (2003) Parent-training programmes for improving maternal psychosocial health, Cochrane Database of Systematic Reviews, 4, Art. No.: CD002020.	Parenting programmes can make a significant contribution to the short-term psychosocial health of mothers
Winokur M, Holtan A, Valentine D (2009) Kinship care for the safety, permanency, and well-being of children removed from the home for maltreatment: review, Campbell Systematic Reviews.	Children in kinship foster care experience better behavioral development, mental health functioning, and placement stability than do children in non-kinship foster care
Cleary M, Hunt GE, Matheson S, Walter G (2009) Psychosocial treatments for people with co–occurring severe mental illness and substance misuse: systematic review, Journal of Advanced Nursing, 65, 2, 238–258.	Main finding supports the use of motivational interviewing in psychiatric settings for the reduction of substance use, at least in the short term.
Cattan M, White M, Bond J, Learmouth A (2005) Preventing social isolation and loneliness among older people: a systematic review of health promotion interventions, Ageing and Society, 25, 1, 41–67.	The review suggests that educational and social activity group interventions that target specific groups can alleviate social isolation and loneliness among older people. The effectiveness of home visiting and befriending schemes remains unclear.

Table 1: (continued)

Study reference	Selected summary findings
Oliver D, Connelly J, Victor C, Shaw F, Whitehead A, Genc Y, et al (2007) Strategies to prevent falls and fractures in hospitals and care homes and effect of cognitive impairment: systematic review and meta–analyses, British Medical Journal, 334, 7584.	There is some evidence that multifaceted interventions in hospital reduce the number of falls and that the use of hip protectors in care homes prevents hip fractures. There is insufficient evidence, however, for the effectiveness of other single interventions in hospitals or care homes or multifaceted interventions in care homes.
Gascoigne C, Morgan K, Gross H, Goodwin J (2010) Reducing the health risks of severe winter weather among older people in the United Kingdom: an evidence-based intervention, Ageing & Society, 30, 2, 275–297.	Systematic review revealed 28 items relevant to reducing mortality, 22 of which were found to be translatable into practice.
Forster A, Lambley R, Hardy J, Young J, Smith J, Green J, et al. (2009) Rehabilitation for older people in long-term care, Cochrane Database of Systematic Reviews, 1, Art. No: CD004294.	Provision of physical rehabilitation interventions to long-term care residents is worthwhile and safe, reducing disability with few adverse events. Most trials reported improvement in physical condition. However, there is insufficient evidence to make recommendations about the best intervention, improvement sustainability and cost-effectiveness.
Braye S, Preston-Shoot M, Cull L, Johns R, Roche J (2005) Learning, Teaching and Assessment of Law in Social Work Education: A Knowledge Review, Knowledge Review 08, Social Care Institute for Excellence, London.	Law education should focus on the circumstances that social workers are likely to encounter but many graduates find it difficult to apply their knowledge to real situations. Most practice programmes do not include formal law-related learning objectives for students on placement, and students struggle to put knowledge into practice.
McNally S, Ben-Shlomo Y, Newman S (1999) The effects of respite care on informal carers' well-being: a systematic review, Journal of Disability and Rehabilitation, 21, 1, 1–14.	There was little evidence that respite intervention has either a consistent or enduring beneficial effect on carers' well-being. The 29 studies included were methodologically poor, but the findings suggest respite care often fails to facilitate the maintenance of socially supportive relationships, which may moderate strain after respite has ended.

Boaz *et al.* (2002) summarise the issues raised in applying traditional review methodology to social science and social care. Traditional methods may misrepresent the relative value of different research methods and undermine the suitability of particular designs for understanding complex interventions and multiple outcomes, and they may exclude theories of change which try to capture some of the intervening stages between intervention and outcome. Social scientists have been influential in increasing user involvement in reviews, developing tools for appraising and systematically synthesising qualitative research, and 'negotiating review questions with stakeholders such as policy makers' (Boaz *et al.* 2002), so as to include perspectives valued by users as well as the technical efficiency of a practice. The increase in systematic reviews in social sciences may also have consequences for the methodological quality of primary research. By and large, we have argued that social scientists have enhanced the traditional role, remit and methodology of systematic reviews.

CONCLUSIONS

Introducing a range of knowledge sources including different types of research study does not displace the systematic review method based traditionally on randomised controlled trials, but it does supplement and improve it, rendering the review more useful and relevant to the users of knowledge. A multi-method approach is also arguably more democratic: the experience of people at the centre of the activity is taken into account (Gough and Elbourne 2002). Combining different types of knowledge within a systematic review is possible and desirable, but requires careful disaggregation and framing of different types of study, and transparency of method and reporting, if the review is to retain the term 'systematic'. The contrast between 'positivist' and 'interpretivist' approaches to research methods and data has become unhelpfully polarised (Gough and Elbourne 2002). We need both quantitative and qualitative research in the social sciences to support policy and practice, and they will be complementary in addressing different aspects of particular topics.

There is general agreement among social scientists that systematic reviews are useful to researchers and decision-makers, but less consensus on what they should include. Extreme positions are taken about the central position of effectiveness studies which employ randomised controlled trial design, and the role of qualitative studies which address the complex contextual factors of the social world, including the views of service users. Increasingly, social scientists seek to develop and implement within review methodology:

- explicit and rigorous appraisal of individual RCTs, as their conduct can significantly bias results;
- more rigorous appraisal, and some standardisation in reporting and appraisal, of qualitative studies;
- examples of studies which have synthesised findings from both formative and summative research, as these offer the best 'coverage' of a topic area.

Better understanding of systematic reviews among the social research community may be facilitated by:

- braver research designs to provide high quality evidence to support reviews (e.g. cluster randomisation);
- more transparency in reporting primary research and secondary analysis;
- more emphasis on knowledge/information science as a specialist skill;
- better understanding of research evidence and research methods;
- better understanding of qualitative synthesis techniques.

As a tool of policy and practice in complex social and technical fields such as social welfare, research reviews cannot tell the whole story. Policymakers and practitioners have to take into account more than research findings, and certainly more than the effectiveness of an intervention. Whatever their limitations, systematic reviews – transparently conducted – remain a valuable means of accessing and understanding what has already been learnt from research.

References

Abrami P, Borokhovski E, Bernard R, Wade CA, Tamim R, Persson T, Bethel E, Hanz K, Surkes M (2010) Issues in conducting and disseminating brief reviews of evidence, *Evidence & Policy*, 6, 3, 371–389.

Attree P, Milton B (2006) Critically appraising qualitative research for systematic reviews: defusing the methodological cluster bombs, *Evidence & Policy: a Journal of Research, Debate and Practice*, 2, 1, 109–126.

Bero LA, Jadad AR (1997) How consumers and policymakers can use systematic reviews for decision making, *Annals of Internal Medicine*, 127, 1, 37–42.

Boaz A, Ashby D, Young K (2002) *Systematic reviews: what have they got to offer evidence based policy and practice?*, ESRC UK Centre for Evidence Based Policy and Practice, London.

Boaz A, Pawson R (2005) The perilous road from evidence to policy: five journeys compared, *Journal of Social Policy*, 34, 2, 175–194.

Boaz A, Solesbury W, Sullivan F (2004) *The practice of research reviewing I. An assessment of 28 review reports*, Working Paper 22 (draft), ESRC UK Centre for Evidence Based Policy and Practice, London.

Booth A (2008) Unpacking your literature search toolbox: on search styles and tactics, *Health Information and Libraries Journal*, 25, 4, 313–317.

Braye S, Preston-Shoot M (2007) On systematic reviews in social work: observations from teaching, learning and assessment of law in social work education, *British Journal of Social Work*, 37, 2, 313–334.

Carr S, Coren E (2007) *Collection of examples of service user and carer participation in systematic reviews*, SCIE Research Resources, Research Resource 02, Social Care Institute for Excellence, London.

Chalmers I, Hedges LV, Cooper H (2002) A brief history of research synthesis, *Evaluation and the Health Professions*, 25, 1, 12–37.

Clapton J, Rutter D, Sharif N (2009) *SCIE Systematic mapping guidance (draft)*, SCIE Research Resource 03, SCIE Research Resources, Social Care Institute for Excellence, London.

Coghlan M, Bergeron C, White K, Sharp C, Morris M, Rutt S (2009) *Narrowing the gap in outcomes for young children through effective practices in the early years*, Centre for Excellence & Outcomes in Children's and Young People's Services (C4EO), London.

Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M (2008) *Developing and evaluating complex interventions: new guidance*, MRC Medical Research Council, London.

Craven P, Lee R (2006) Therapeutic interventions for foster children: a systematic research synthesis, *Research on Social Work Practice*, 16, 3, 287–304.

Davies P, Boruch RF (2001) The Campbell Collaboration does for public policy what Cochrane does for health, *British Medical Journal*, 323, 294–295.

Department of Health (2006) *Our health, our care, our say: a new direction for community services*, Cm 6737, Department of Health, London.

Dickson K, Gough D (2009) *Supporting people in accessing meaningful work: recovery approaches in communitybased adult mental health services*, SCIE Knowledge Review 21, SCIE, London.

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

Dixon-Woods M, Cavers D, Agarwal S, Annandale E, Arthur A, Harvey J (2006) Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups, *BMC Medical Research Methodology*, 6, 1, 35.

Dixon-Woods M, Sutton A, Shaw R, Miller T, Smith J, Young B, et al. (2007) Appraising qualitative research for inclusion in systematic reviews: a quantitative and qualitative comparison of three methods, *Journal of Health Services Research and Policy*, 12, 42–47.

Doll P, Peto R, Boreham J, Sutherland I (2008) Mortality in relation to smoking: 50 years' observations on male British doctors, *British Medical Journal*, 328, 1519, doi:10.1136/bmj.38142.554479.

Edwards A, Elwyn G, Hood K, Rollnick S (2000) Judging the 'weighting of evidence' in systematic reviews: introducing rigour into the qualitative overview stage by assessing signal and noise *Journal of Evaluation in Clinical Practice*, 6, 177–184.

EPPI-Centre (Sept 2006) *EPPI-Centre methods for conducting systematic reviews* EPPI-Centre, Social Science Research Unit, Institute of Education, University of London, London.

Fisher M (2002) The Social Care Institute for Excellence: the role of a national institute in developing knowledge and practice in social care, *Social Work & Social Sciences Review*, 10, 2, 6–34.

Fisher M (2005) Knowledge production for social welfare: enhancing the evidence base, in Sommerfeld P (ed.) *Evidence-Based Social Work. Towards a new professionalism?*, 127–147, Peter Lang, Bern.

Fisher M, Qureshi H, Hardyman W, Homewood J (2006) *Using qualitative research in systematic reviews: Older people's views of hospital discharge*, SCIE, London.

Glasby J, Beresford P (2006) Commentary and issues: who knows best? Evidence-based practice and the service user contribution, *Critical Social Policy*, 26, 1, 268–284.

Gough D, Elbourne D (2002) Systematic research synthesis to inform policy, practice and democratic debate, *Social Policy and Society*, 1, 3, 225–236.

Gough DA, Kiwan D, Sutcliffe S, Simpson D, Houghton N (2003) *A systematic map and synthesis review of the effectiveness of personal development planning for improving student learning*, EPPI-Centre, Social Science Research Unit, Institute of Education, University of London, London.

Greenhalgh T, Kristjansson E, Robinson V (2007) Realist review to understand the efficacy of school feeding programmes, *British Medical Journal*, 335, 858–861.

Greenhalgh T, Peacock R (2005) Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources, *British Medical Journal*, bmj.38636.593461.593468.

Grégoire G, Derderian F, Le Lorier J (1995) Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias?, *Journal of Clinical Epidemiology*, 48, 1, 159–163.

Guyatt G, Sackett D, Sinclair J, Hayward R, Cook D, Cook R (1995) Users' guide to the medical literature IX: A method for grading health recommendations, *Journal of the American Medical Association*, 274, 1800–1804.

Hammersley M (2001) On 'systematic' reviews of research literatures: a 'narrative' response to Evans & Benefield, *British Educational Research Journal*, 27, 5, 543–554.

Hansen HF, Rieper O (2009) The evidence movement: the development and consequences of methodologies in review practices, *Evaluation*, 15, 2, 141–163.

Harden A, Garcia J, Oliver S, Rees R, Shepherd J, Brunton G, Oakley A (2004) Applying systematic review methods to studies of people's views: an example from public health research, *Journal of Epidemiology & Community Health*, 58, 9, 794–800.

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

Higgins JPT, Green S (eds) (2006) *Cochrane Handbook for Systematic Reviews of Interventions*, John Wiley & Sons, Chichester.

Kuepper R, van Os J, Lieb R, Wittchen H, Höfler M, Henquet C (2011) Continued cannabis use and risk of incidence and persistence of psychotic symptoms: 10 year follow–up cohort study, *British Medical Journal*, 342:d738.

Lavis J, Davies H, Oxman A, Denis J-L, Goen-Biddle K, Ferlie E (2005) Towards systematic reviews that inform health care management and policy-making, *Journal of Health Services Research & Policy*, 10, 3 Supplement, 35–48.

Lees J, Manning N, Rawlings B (1999) *Effectiveness of Therapeutic Communities: A systematic international review of therapeutic community treatment for people with personality disorder and mentally disordered offenders*, CRD Report 17, NHS Centre for Reviews and Dissemination, University of York, York.

Lewin G, Vandermeulen S (2009) A non-randomised controlled trial of the Home Independence Program (HIP): an Australian restorative programme for older home-care clients, *Health & Social Care in the Community*, 18, 1, 91–99.

Lieberman HM, Hunt IF, Coulson AH, Clark VA, Swendseid ME, Ho L (1976) Evaluation of a ghetto school breakfast program, *Journal of American Diet Assoication*, 68, 132–138.

Littel JH, Corcoran J, Pillai V (2008) *Systematic Reviews and Meta-Analysis*, Oxford University Press, Oxford.

Lloyd Jones M (2004) Application of systematic review methods to qualitative research: practical issues, *Journal of Advanced Nursing*, 48, 3, 271–278.

Marsh P, Fisher M (2005) *Developing the evidence base for social work and social care practice*, SCIE Report 10, Social Care Institute for Excellence, London.

Marshall M, Lockwood A (1998) Assertive community treatment for people with severe mental disorders, *Cochrane Database of Systematic Reviews*, Issue 2, Art No. CD001089. DOI: 10.1002/14651858.CD001089.

Mason A, Weatherly H, Spilsbury K, Arksey H, Golder S, Adamson J, Drummond M, Glendinning C (2007) *A systematic review of the effectiveness and cost-effectiveness of different models of community–based respite care for frail older people and their carers*, Health Technology Assessment, 11, National Coordinating Centre for Health Technology Assessment, Southampton.

Mays N, Pope C (2000) Qualitative research in health care: assessing quality in qualitative research, *British Medical Journal*, 320, 50–52.

Moher D, Fortin P, Jadad A, Juni P, Klassen T, Le Lorier J, Liberati A, Linde K, Penna A (1996) Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of sytematic reviews, *Lancet*, 347, 8998, 363–366.

Moher D, Schulz K, Altman D (2001) The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials, *BMC Medical Research Methodology*, 1, 2.

Mullen EJ (2006) Choosing outcome measures in systematic reviews: critical challenges, *Research on Social Work Practice*, 16, 1, 84–90.

National Institute for Health and Clinical Excellence (2009) *Methods for the development of NICE public health guidance (second edition)*, National Institute for Health and Clinical Excellence, London.

NHS Centre for Reviews and Dissemination (2001) *Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commisioning reviews*, CRD Report Number 4 (2nd Edition), Khan KS, ter Riet G, Glanville J, Sowden AJ, Kleijnen J, NHS Centre for Reviews and Dissemination, University of York, York.

Noblitt GW, Hare DW (1988) *Meta-Ethnography: Synthesizing Qualitative Studies*, Sage, Beverley Hills.

Noyes J, Popay J, Pearson A, Hannes K, Booth A, Cochrane Qualitative Methods Group (2008) Chapter 20: Qualitative research and Cochrane Reviews, in Higgins JPT, Green S (eds) *Cochrane Handbook for Systematic Reviews of Interventions*, Version 5.0.1, The Cochrane Collaboration.

Oakley A (2003) Research evidence, knowledge management and educational practice: early lessons from a systematic approach, *London Review of Education*, 1, 1, 21–33.

Oakley A, Gough D, Oliver S, Thomas J (2005) The politics of evidence and methodology: lessons from the EPPI-Centre, *Evidence & Policy: a Journal of Research, Debate and Practice*, 1, 1, 5–31.

Olsson T (2007) Reconstructing evidence-based practice: an investigation of three conceptualisations of EBP *Evidence & Policy: a Journal of Research, Debate and Practice*, 3, 2, 271–85.

Parker G, Bernard S, Gridley K, Aspinal F, Light K (2010) *Rapid Systematic Review of International Evidence on Integrated Models of Care for People with Long-term Neurological Conditions. Technical Report*, SPRU, University of York, York.

Pawson R (2001) *Evidence Based Policy: II. The promise of 'realist synthesis'*, Working Paper, ESRC Network for Evidence Based Policy and Practice, London.

Pawson R (2002) *Does Megan's Law Work? A theory-driven systematic review*, Working Paper 8, ESRC UK Centre for Evidence Based Policy & Practice.

Pawson R (2004) *Assessing the quality of evidence in evidence-based policy: why, how and when?*, ESRC Research Methods Programme Working Paper No 1, University of Manchester, Mancester.

Pawson R, Boaz A, Grayson L, Long A, Barnes C (2003) *Types and quality of social care knowledge*, SCIE knowledge review 3, SCIE, London.

Pawson R, Greenhalgh T, Harvey G, Walshe K (2005) Realist review – a new method of systematic review designed for complex policy interventions, *Journal of Health Services Research & Policy*, 10, 3 Supplement, 21–34.

Pearson A (2004) Balancing the evidence: incorporating the synthesis of qualitative data into systematic reviews, *JBI Reports*, 2, 2, 45–64.

Petrosino A, Turpin-Petrosino C, Buehler J (2003) *'Scared Straight' and other juvenile awareness programs for preventing juvenile delinquency (updated review)*, Campbell Collaboration Systematic Reviews, 2004.2.

Petticrew M (2003) Why certain systematic reviews reach uncertain conclusions, *British Medical Journal*, 326, 756–758.

Petticrew M, Roberts H (2006) *Systematic Reviews in the Social Sciences: a Practical Guide*, Blackwell, Oxford.

Pignotti M, Mercer J (2007) Holding Therapy and Dyadic Developmental Psychotherapy are not supported and acceptable social work interventions: A systematic research synthesis revisited, *Research on Social Work Practice*, 17, 513–519.

Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, Britten N (2006) *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews*, ESRC Research Methods Programme, University of Lancaster, Lancaster.

Pope C, Mays N (1993) Opening the black box: an encounter in the corridors of health services research, *British Medical Journal*, 306, 315–318.

Reilly S, Xie C, Jacobs S, Challis D (2008) Examining the state of adult social care research 1990–2001: a systematic synthesis of research methods and quality, *Evidence & Policy: a Journal of Research, Debate and Practice*, 4, 3, 155–182.

Roen K, Arai L, Roberts H, Popay J (2006) Extending systematic reviews to include evidence on implementation: Methodological work on a review of community-based initiatives to prevent injuries, *Social Science & Medicine*, 63, 1060–71.

Rutter D, Fisher M, Francis J (2010) *SCIE systematic reviews: Guidelines (2nd edition)*, Research Resource 01, SCIE, London.

Sandelowski M, Barroso J (2007) *Handbook for synthesizing qualitative research*, Springer Publishing Company, New York.

Slavin RE (1995) Best evidence synthesis: an intelligent alternative to meta-analysis, *Journal of Clinical Epidemiology*, 48, 1, 19–21.

Smedlund G, Dalsbø T, Steiro A, Winsvold A, Clench-Aas J (2007) *Cognitive behavioural therapy for men who physically abuse their female partner*, Cochrane Collaboration.

Spencer L, Ritchie J, Lewis J, Dillon L (2003) *Quality in qualitative evaluation: a framework for assessing research evidence*, Government Chief Social Researcher's Office, London.

Taylor BJ, Dempster M, Donnelly M (2007) Grading gems: appraising the quality of research for social work and social care, *British Journal of Social Work*, 37, 2, 335–354.

Wade CA, Turner HM, Rothstein HR, Lavenberg JG (2006) Information retrieval and the role of the information specialist in producing high-quality systematic reviews in the social, behavioural and education sciences, *Evidence & Policy: a Journal of Research, Debate and Practice*, 2, 1, 89–108.

Weaver T, Hart J, Rutter D, Metrebian N, Chantler K (2007) *National Evaluation of Crack Cocaine Treatment & Outcome Study (NECTOS): A Multi-Centre Evaluation of Dedicated Crack Cocaine Treatment Services*, National Treatment Agency, London.

Zwi K, Woolfenden S, Wheeler D, O'Brien T, Tait P, Williams K (2007) *School-based education programmes for the prevention of child sexual abuse*, Cochrane Collaboration.

Further references are given in full in Table 1.