2014, 1–17, Early Online MEDICAL TEACHER

AMEE GUIDE

Systematic reviews in medical education: A practical approach: AMEE Guide 94

RICHA SHARMA[1], MORRIS GORDON[2], SHAFIK DHARAMSI[1] & TREVOR GIBBS[3]
[1]University of British Columbia, Canada, [2]University of Central Lancashire, UK, [3]AMEE, UK

Abstract

The twentieth century saw a paradigm shift in medical education, with acceptance that 'knowledge' and 'truth' are contextual, in flux and always evolving. The twenty-first century has seen a greater explosion in computer technology leading to a massive increase in information and an ease of availability, both offering great potential to future research. However, for many decades, there have been voices within the health care system raising an alarm at the lack of evidence to support widespread clinical practice; from these voices, the concept of and need for evidence-based health-care has grown. Parallel to this development has been the emergence of evidence-based medical education; if healthcare is evidence-based, then the training of practitioners who provide this healthcare must equally be evidence-based. Evidence-based medical education involves the systematic collection, synthesis and application of all available evidence, when available, and not just the opinion of experts. This represented a seismic shift from a position of expert based consensus guidance to evidence led guidance for evolving clinical knowledge. The aim of this guide is to provide a practical approach to the development and application of a systematic review in medical education; a valid method used in this guide to seek and substantiate the effects of interventions in medical education.

Introduction

The origins of medical education were grounded in the practice of apprenticeship as long as two millennia ago, with knowledge viewed as a commodity to be delivered directly to the learner (Drabkin 1957). This knowledge could develop as expertise, but essentially was seen as 'truth' to be transmitted to learners. The twentieth century saw a paradigm shift in this viewpoint, with acceptance that 'knowledge' and 'truth' are contextual, in flux and always evolving (Sackett et al. 1996). The technology explosion has led to a general ease of access to the massive increase in information, not only offering great potential but also inordinate risk (Altman 1994). The most prominent concern raised by doctors since the outset of this revolution has been the poor quality of much of the available information (Shactman 2000) and for many decades, there have been voices within health care raising alarm at the lack of evidence to support widespread clinical practice (Mulrow 1987; Sackett & Rosenberg 1995) The thousands of irrelevant studies that appear using an online search has led to the 'fool's gold of the digital age' (Gordon et al. 2013a). There is an even greater challenge in the field of medical education, where multiple research methodologies are used by scholars from ideologically polarised backgrounds to answer the same question (Cresswell et al. 2010). All this has led to the evidence-based medicine movement, which originated at the McMaster University in Canada where it was defined as 'the conscientious, explicit and judicial use of current evidence in making decisions about the care of individual patients' (Sackett et al. 1996).

Practice points

- Evidence to support, refute or guide effective medical education increases at an inexorable rate.
- Much of the data describing the newer approaches to healthcare and the education of doctors are not always bound in evidence or validity, and frequently the findings are not always transferable to other educational situations.
- There is a need to have a logical approach to gathering the educational research data and setting it out so that it becomes an effective tool in the educational researcher's toolbox.
- A medical education systematic review is an effective research tool but requires a dedicated approach, based upon a series of accepted steps to development.

Evidence-based health care involves the systematic collection, synthesis and application of all available evidence, when available, not just the opinion of experts (Moher et al. 1999). This represented a seismic shift from a position of expert-based consensus guidance to evidence led guidance for evolving clinical knowledge (Burgers et al. 2003). The most important element of the evidence-based health care movement is an acceptance of the evolving nature of 'truth'. Researchers have sought to quantify this, no more elegantly than Hall and Platell (1997). They demonstrated that the half-life of clinical truth in the surgical field is 45 years and therefore

Correspondence: Dr. Morris Gordon, School of Medicine and Dentistry, Harrington Building Room HA106, University of Central Lancashire, Fylde Road, Preston, Lancashire, PR1 2HE, UK. Tel: +44 (0)1253 303443; E-mail: morris@betterprescribing.com

ISSN 0142-159X print/ISSN 1466-187X online/14/000001–17 © 2014 Informa UK Ltd.
DOI: 10.3109/0142159X.2014.970996

within half a century, 50% of what is known is no longer accurate. This more than anything cements the need for a contemporaneous and evidence informed knowledge base, rather than an expert led knowledge base (Poynard et al. 2002).

The cochrane collaboration

The strength of systematic review was especially demonstrated in a key appraisal describing the efficacy of corticosteroids given to pregnant women who deliver premature babies (Crowley et al. 1990). The results revealed that administration of maternal corticosteroids significantly reduced morbidity and mortality among premature infants. The celebration of this discovery was tempered however by the realisation that a similar meta-analysis of data up to a decade earlier in 1980 showed the same result. So, uneasy were the combined impact of the missed data and the realisation of fact, that it inspired the formation of one of the key entities in the globe in the field of evidence-based health care, the Cochrane Collaboration (2013).

The Cochrane Collaboration is a global movement with the key objectives of preparing, maintaining and disseminating results of systematic reviews of health care interventions. Cochrane led the formulation of the systematic approach to evidence synthesis, as categorised by systematic review (Doshi et al. 2012), to deal with the issues already highlighted by misuse of the tools of evidence-based health care (Moher et al. 1999). This included writing a concise review protocol that is reviewed prior to the commencement of work and the use of clear criteria regarding inclusion and exclusion, quality, strength of conclusions and lay summaries. Cochrane reviews are viewed as the benchmark in supporting evidence based decision-making (Olsen et al. 2001). Similar organisations developed symbiotically through the last 20 years, including the Campbell collaboration focussing on education and justice (Campbell Collaboration 2013), as well as EPPI centre in public health and education policy (Social Science Research Unit 2009).

Medical education evidence synthesis

In the world of medical education, the issues of evidence synthesis are far more complex and challenging. For over a decade, there have been calls for medical education to become more evidence-based (Bligh 2000; Carline 2004; Chen et al. 2005). An article in the British Medical Journal (Todres et al. 2007) sparked an active debate regarding the nature of quality within medical research, a key issue when synthesising evidence. Scholars in the field recognised that the rich tapestry of research types used in education does not invalidate the issue of quality (Gordon et al. 2013a), but merely indicates that measures used in clinical medicine (Gutiérrez Castrellón et al. 2010) may not be appropriate to measure quality in this context (Norman 2003). This has meant that developing such systematic review approaches in medical education present new and unique opportunities as well as challenges.

BEME collaboration

The best evidence medical education (BEME) collaboration was established in 1999 (Harden et al. 1999) as an effort to move the use of anecdotal information in medical education to the use of evidence synthesis through systematic review, mirroring the revolution seen in healthcare. Their mission statement is '... disseminating information which allows medical teachers, institutions and all concerned with medical education to make decisions on the basis of the best evidence available; producing appropriate systematic reviews of medical education which reflect the best evidence available and meet the needs of the user; and creating a culture of best evidence medical education amongst individual teachers, institutions and national bodies' (www.bemecollaboration.org).

In completing these goals, the first challenge that BEME had to overcome was to recognise the unique challenges of evidence synthesis in this field and to produce tools to support authors in using a clear and reliable methodology. In line with the work of the Cochrane Collaboration, the methodology that BEME champions involves a comprehensive search of all potentially relevant studies and the use of explicit, reproducible criteria in the selection of studies for review (Patricio & Vaz Careneiro 2012). In achieving this goal, they attempted to grapple with the concept of evidence synthesis methodology within the very complex context of health education, producing often revised guidance pieces for researchers (Hammick et al. 2010). These works have provided insight into these key methodological issues when establishing the process of systematic review in the context of medical education, such as sources of medical education evidence (Haig & Dozier 2003a) and how to construct a search of these evidence sources (Haig & Dozier 2003b). BEME has led the way in this area and their activities have contributed significantly to practice and essentially founded the process of evidence synthesis in health education.

However, whilst BEME endeavours to be extremely inclusive, supportive and accessible, not all authors will choose to complete their systematic review within the BEME collaboration. Within the wider field, such publications can be of variable quality and value to readers (Gordon & Gibbs 2014). Attempts have been made recently to produce explicit guidance on the reporting of such systematic reviews (Gordon & Gibbs 2014). However, it is outside of the scope of those works to offer explicit methodological guidance.

Characteristics of a systematic review

A systematic review responds to a specific research question that is relatively narrow in scope: e.g. what impact do structured educational sessions to increase emotional intelligence have on medical students? (Cherry et al. 2012) – and provides an in-depth analysis and summary of the best available evidence in response to this question. This involves a rigorous process of searching, selecting, appraising, interpreting and summarising results from published studies on this specific topic (Crowther & Cook 2007). It is an important tool for professionals who seek the best available evidence to inform their actions (EBBP 2013). The original studies

reviewed in the education setting are as varied as in healthcare itself – they can be observational or experimental, quantitative or qualitative. A review can be described as systematic (EBBP 2013; Khan et al. 2003), if it has:

- clearly formulated question(s),
- identifies relevant studies,
- critically appraises their quality,
- summarises the evidence using an explicit methodology.

Systematic reviews must contribute to the literature by filling a gap in published reviews, and adding significantly to the current body of knowledge in terms of quality of data (Cook & West 2012). Done well, systematic reviews can be used for making recommendations and developing guidelines, informing public policy, assessing performance measures, setting research agendas, making decisions regarding individual patient care or teaching practice and aiding in decision-making (EBBP 2013). It is also possible to conduct a systematic review of systematic reviews (Smith et al. 2011).

Contrasting with a literature review

Literature reviews are common within the peer-reviewed literature in all areas of academia. The authors often identify such reviews as a brief overview with no specific review question (Moher et al. 2007). The topic of examination is often quite broad, and the sources of literature are not necessarily specified; there are no attempts to obtain everything that is written on the topic, and whilst the papers reviewed are summarised and critiqued, this is not usually done in a rigorous or a systematic fashion (Magarey 2001). The research is further summarised in a narrative style, making it susceptible to bias where the reviewer might only select articles that support their views, preventing the presentation of a broader view of the research. Promoting eminence over evidence, this type of non-systematic authoritative review is not suitable for generating evidence regarding the effectiveness of particular interventions (Magarey 2001).

It is the rigor and methodology of the systematic review process that differentiates it from the traditional literature review and so can be considered as a secondary form of research in its own right (Magarey 2001; EBBP 2013). Alternatively, we can think of a systematic review as a literature review conducted with explicit, rigorous and transparent methods (Greenhalgh et al. 2005). In a study examining the attitudes of editors of core clinical journals towards systematic reviews and their value for publication, it was found that most editors consider them to be original research (Meerpohl et al. 2012). The characteristics of literature reviews, compared to systematic reviews are seen in Table 1.

Why the need for this guide

The BEME collaboration continues its work to champion the high quality and utility of systematic review within evidence synthesis, publishing what many consider the gold standard evidence synthesis in the field. However, the field is rapidly expanding, with most reviews still published outside of the BEME collaboration. As systematic reviews in education become increasingly common, journal editors and chairs of review committees are noting that many applicants fail to address key actions required in a rigorous review (Cook & West 2012) and this ultimately leads to final publications that are limited in their value to readers (Gordon et al. 2014). Whilst recent work has offered clear guidance on the reporting of such works (Gordon & Gibbs 2014), there is a need for detailed guidance on how to plan and conduct high-quality systematic reviews in the field of medical education.

When medical educators are equipped with limited resources, a systematic review can be an invaluable tool to inform and create the most effective interventions and policies (Yousefi-Nooraie 2009; Mbuagbaw et al. 2011). A systematic review can help identify further areas of research, as well as where it might be unnecessary or even unethical (Chalmers & Glasziou 2009). Thus, funding agencies such as the UK Medical Research Council require systematic reviews to be part of grant applications (Meerpohl et al. 2012). Leading medical journals also advocate for systematic overview of the evidence as part of reports of new randomised trials (Clarke et al. 2010), and in 2009, the US government allocated $1.1 billion to comparative-effectiveness research (Meerpohl et al. 2012).

Despite the support from funding agencies and government institutions, the value of systematic reviews is still debated within academia. Whilst some consider it to be 'secondary research', others argue that it comprises to be original research, based on its rigorous scientific methodology. Recent publications have highlighted the potential for the systematic review of primary educational material to support generation of new knowledge (Hammick et al. 2010; Gordon et al. 2013a), further cementing the future potential of these techniques in the health education field.

Table 1. Characteristics of systematic and literature reviews.

Literature reviews	Systematic reviews
A brief overview with no specific research question	Clear, explicit objectives with stated inclusion criteria for studies to be selected
Sources of literature and selection process of studies are not specified in depth	Use systematic search methods reducing sampling bias
Papers reviewed are summarised and critiqued, but not done in a rigorous or a systematic fashion	Use consistent evaluation of available information such as outcomes and study quality
Susceptible to bias where the reviewer might only select articles that support their views, preventing the presentation of an objective view of the research	Increased transparency showing how decisions were made in the review process, enabling direct assessment of review quality
Not considered as original research	Can be considered original research, but often called secondary research
Integrate research from diverse fields and identify new insights	Summarises research on a focused topic and highlight its strengths and weaknesses in the existing body of evidence

The defining characteristics of a systematic review are clear, but in the context of health education, there are a number of different forms of review that may exist that can lead to confusion in nomenclature. For example, in recent years, there has been great interest in realist reviews (Wong et al. 2013), but can these be considered a form of systematic review? The answer is the same as for all reviews – they *can* be if they meet all the criteria that is described above. However, a realist synthesis can also be carried out in a non-systematic manner and as such, clear reporting to support readers in making judgements are required.

In this guide, we will present the various steps in completing a medical education systematic review outside of the supporting structures of an organisation such as BEME. It is not possible to offer an exhaustive digest on the wide range of methodologies available to reviewers. Our aim however, is to present an overview that can help guide planning and decision making for all such works, focussing on the homogenous elements of such works, as well as directing readers to further sources of information.

Conducting a health education systematic review

Conducting a systematic review is a complex undertaking; many consider it a research project in itself. Researchers undertaking this task must realise that a well-conducted review is a time intensive project rather than a 'quick and easy' task (Cook & West 2012). Paradoxically, the process itself is not only innately difficult or challenging, but also involves following a logical process in a step-wise fashion. As such, the techniques are fairly inclusive and can form a very useful first step for those who have not previously been involved in formal research activities.

Step 1: Inception of review

The first phase of a medical education systematic review is best considered from the perspective of a problem. Identifying such a problem can help guide the form, method, scope and focus of a review. This is best illustrated from the perspective of an example – *teaching evidence based medicine techniques*, for instance. This may seem a sensible topic for a review. However, now consider this topic from the perspective of several problems.

(a) A Head of Department in an undergraduate medical school believes that such education (teaching evidence-based medicine) cannot change the behaviour of students to make them more likely to employ evidence. You therefore decide to perform a systematic review to establish whether evidence-based medicine courses for medical students can impact their use of such techniques in practice.

(b) When you present this idea to your Head of Department she is impressed. However, she finds such a review published last month. Instead, she suggests it would be useful to establish learning outcomes and content. You therefore seek to perform a systematic review to establish what the learning outcomes are, teaching methods and content to teach evidence-based medicine that should be used in undergraduate medical courses.

(c) A nursing colleague who is interested in educational research likes this idea, but proposes that an additional aspect that needs to be considered is not whether or what evidence based medicine teaching is effective, but actually how and why such education may change behaviour. He is aware that such research has been performed and so you instead refocus your review. You decide to investigate how and why evidence-based medicine teaching changes health professional behaviour.

These examples demonstrate that what essentially seems like the same basic topic can in fact inspire numerous different systematic reviews, all with their own merit, outcomes, scope and requirements. Whilst these examples are presented to illustrate a point, it is suggested that the first stage of a review should be to carefully consider, dissect and define a problem that is important to your local setting, region or the wider educational or research community.

It cannot be over-emphasised the importance of this stage; clarity at this time frequently leads to clarity and ease of progression in the later stages.

Step 2: Scoping searches

When you have a clear idea of a problem and a succinct question arising, and with a resulting topic of a review, the next step is to approach the existing evidence base in that field. Clearly, such a search will be by its nature a non-systematic review and as such, party to the biases and other difficulties already identified. This can be done by checking the Cochrane database of systematic reviews, previous BEME reviews, the Database of Abstracts of Reviews of Effectiveness and other electronic databases, such as Pubmed (Magarey 2001).

The purpose of the scoping review is two-fold. First, to identify any existing evidence synthesis works that have been completed; these may make the review to be undertaken redundant, identify an area for refocus or even strengthen the need for the review. Second, increased awareness of the breadth and depth of the existing evidence can support the next phases of the project. This knowledge can inform team selection, inform selection of appropriate analytical techniques and guide appropriate specific research objectives.

Step 3: Assembling the full review team

Systematic reviews are a team activity, and choosing the right team members is an essential part of the entire review process. Step 1 clearly will inform this part of the process, but resource limitations and other pragmatic considerations may come into play. Whilst the rigor of the review cannot be compromised due to such considerations, they can inform and allow for refinement of an idea to reduce the size of the project. Diversity of perspectives helps to enrich discussions and enhances the quality and generalisability of the review.

When assembling the team, one should also consider covering other areas of expertise (EBBP 2013). For instance, a strong team would include individuals who are:

- *Experienced on systematic review methods*: One or more persons in the team should know how to conduct a systematic review. This person can guide the development of procedures and protocols, as well as supporting educational development of other members of the team;
- *Familiar with the content of the review*: One or more persons in the team should provide expertise on the topic of the area. In general, diverse perspectives from practitioners and researchers are useful;
- *Methodological expert*: Depending on the specific techniques to be employed, individuals with experience in various qualitative methodologies may be needed;
- *Statistician*: If conducting a meta-analysis, a statistician familiar with the methodology is necessary;
- *Medical Librarian*: A librarian familiar with searching and documentation procedures of a systematic review is needed in order to conduct rigorous database searching, compiling the body of evidence. Whilst not vital as a core team member, access to the expertise of such a professional is useful; and
- *Data management*: A systematic review covers thousands of abstracts, and a person must be responsible for maintaining the database of references, as well as tracking the status of each abstract (included, excluded etc.) and finally supporting analysis using appropriate software packages.

The team should discuss the expected workload and scope of the project early on in the process to avoid delays in the research plan (Cook & West 2012).

Step 4: Creating the protocol (work-plan)

One of the main ways a systematic review differs from a traditional review is in the requirement for the creation of a protocol. This prospectively lays out a clear plan of action for the review, considering issues of rationale, methodology and scope in a thorough manner. Since a systematic review is a methodical exercise, a protocol should be developed and the process must follow a standard sequence of procedure (Crowther & Cook 2007). Writing a project protocol is a crucial element that provides rigour and guidance during the process (Cook & West 2012). The protocol must be written immediately after or during the writing of the research question in as much detail as possible (EBBP 2013).

A protocol incorporates specific plans for each of the elements of a successful systematic review and is the key characteristic that ensures the quality of the systematic review process. It may be revised as the project progresses and more is learned about the study question, but the ability to refer to a core protocol during the review process allows modifications to be tracked and allows for reproducibility of all steps in the review (Cook & West 2012).

Step 5: Formulating the review question

Part of developing a protocol also involves creating a precise and focused review question; this is a much harder task than expected. It usually requires discussions with collaborators and undergoes many drafts before reaching its final form. The importance of formulating the research question cannot be under-estimated, as it will establish the framework for every following step (Cook & West 2012). It is also important to have no prior expectation of a positive result as this can bias the review (Magarey 2001).

Within clinical medicine, the PICO mnemonic is often used in formulating the review question (Crowther & Cook 2007; Bath & Gray 2009; EBBP 2013) and can similarly be used within medical education evidence synthesis:

Population; **I**ntervention (exposure); **C**omparative interventions; **O**utcomes of interest. For example, **P**: In undergraduate medical students, can, **I**: Problem Based learning, **C**: in comparison to traditional didactic teaching, **O**: enhance team-working skills?

Even though a question can be stated clearly and concisely using the PICO format, the nature of such projects within medical education frequently means a different approach is needed. Indeed, a question is best framed in the context of the current knowledge base and the needs of the field. As such, we would suggest the following **CAPS** format: **C**urrent state of knowledge, **A**rea of interest, **P**otential impact for education and **S**uggestions from experts in the field

***C**urrent state of knowledge*. The scoping in Step 2 should have indicated the current state of knowledge, but further searching may be needed to start to refine the effect of various factors, such as context, learning, pedagogy, etc. The searches may also highlight a lack of research in a given area. The team must then consider whether to interpret this as a need for a detailed and robust search to confirm the limitations of the field or in fact consider that a review in this area may be futile and do little to inform, given the lack of information.

***A**rea of interest*. The information uncovered by the searches above will in many situations lead the authors to a variety of highlighted choices: Should the focus of a review be on confirming the effectiveness or utility of teaching or assessments (justification), on summarising the characteristics of a given element (descriptive) or finally on synthesising evidence to provide new theoretical or conceptual knowledge (clarification)? (Cook et al. 2008). All of these questions may be relevant and it is possible to address more than one, but each element requires different methodological choices to be made and will lead to a different outcome for readers (see below).

***P**otential impact for education*. This is intrinsically linked to the item above, and a vital part of any review; indeed it is a vital element in any research. What impact will a review addressing each of the areas of interest have for educators in the field? In particular, consideration of whether a review focusing on justification of education is of benefit to the field is strongly suggested. Given the intrinsic difficulties in comparing and assessing the effectiveness of different forms of education, as well as the limited use of such findings in pushing forward the field for educators (Gordon 2014), it is a key to consider

how the resulting report will be of use to readers and frame the research question to achieve these goals.

***S**uggestions from experts in the field*. The final element that can be used to support the framing of a question is to consider the opinions of experts in the field. This may be from conference proceedings, commentary and editorial pieces, policy documents and national reports. All will reflect expert opinion and in particular help to identify need, which in turn should direct formal questions for the review.

From these discussions, the reader will realise that these early steps are frequently iterative and observations or findings in a later stage may result in the team looking back and perhaps making adjustments, additions, extractions or other changes to earlier stages.

Step 6: Planning the search

There are a number of key steps in planning the practical search strategy, but it is important to always be repetitively mindful of the question that has been formulated (Cook & West 2012). A search can be systematic and methodologically sound, but if it does not reveal the information that will answer the question of the review, it has been used in vain. In addition, the question of efficacy should also be considered; adding extra steps that add considerable amounts of work with no further yield of studies is neither beneficial nor encouraging for the team.

Whilst some of the operational definitions for searching will be defined from the outset, many others will emerge during the pilot search (below) and during the review process. When the reviewers come across publications, they are uncertain about regarding the inclusion criteria, a rule should be formulated that can aid with decision-making and be applicable in such circumstances. These decisions, along with examples of what should and should not be included must be catalogued in the protocol. Keep in mind that although the conceptual definitions should remain unchanged, the protocol and the operational definitions it contains will continue to evolve during the review process. The development of these items should be a team exercise, determined by all the reviewers in the group. Doing so will not only increase the likelihood that others will agree with the decisions made, but also ensure that everyone applies the criteria in their search, based on a shared understanding. After each round of pilot testing, all reviewers can compare their decisions and refine the operational definitions to maintain consistency.

There is no threshold number of sources for searching that equates to a high quality search. Indeed, often adding more and more electronic databases increases the number of citations, but has no impact on the number of relevant studies elicited. Designing a search strategy should use the pilot search that seeks to consider the relationship between total citations and relevant citations. Considering the studies that have already been found in the scoping search is a useful way of checking the validity of the search – clearly, if it fails to find such relevant studies, the search strategy must be revised.

Inclusion/exclusion criteria

Using the clear question formulated in the last step, a very clear set of inclusion and exclusion criteria must be developed. This step is very important as it impacts upon the generalisability of the results. Decisions on what to include/exclude should be made on a conceptual basis as opposed to convenience (Cook & West 2012). Resource limitations will be a consideration and certain decisions may have to be made that limit the scope of a search to ensure its feasibility. As long as such decisions are clearly and transparently presented to readers, this is not a significant methodological problem.

The criteria to include or exclude articles usually emerge naturally from the focused question and the PICO or CAPS framework (EBBP 2013). For instance, some factors to consider include:

Population. What is the age range of the population? What learning environments are being considered? What professional groups will be included/excluded? Are there any geographical limits? e.g. Undergraduate student health visitors in full time education in the USA.

Intervention. How will a certain intervention be defined? What are the key characteristics that must be demonstrated for inclusion/exclusion? Authors may use formal definitions from dictionaries, theories or previous reviews. Detailed explanations and elaborations can help reviewers recognise key concepts reported in the publications, but utility should be considered to ensure swift and consistent judgments for inclusion can be made. e.g. Educational interventions that are face-to-face and involve a facilitator. Virtual/online/pre-recorded or self-directed courses will all be excluded.

Comparison. Similar questions must be asked of any comparison.

Outcome. This is linked to the outcomes of the CAPS process. Are outcomes that focus on description, justification or clarification to be considered? Are there any recognised outcome measures that must be used for inclusion? e.g. Studies that describe an educational intervention of any kind will be considered (if details are not presented, the authors will be contacted for further information – if there is no response, the study will be rejected)

There are other points to consider as well regarding the study design and biases (EBBP 2013). Sometimes researchers exclude articles based on:

- *Language*: whether to exclude non-English publications;
- *Publication date*: will there be any limits on the age of studies;
- *Rigour or peer review*: excluding graduate theses, presentation papers, unpublished works and other grey literature;
- *Conference proceedings*: This is a contentious issue, as there is some evidence that presentations vary extensively from peer reviewed publications (Pitkin et al. 1999) and therefore must be used with caution. However, as there is often a significant delay between the completion of research and publication in a peer reviewed journal, ignoring such works risks the review being out of date immediately on completion. A balanced approach may consider such works, but only if the author can provide

further methodological and outcome data on request, thus enhancing the quality and usefulness of the work to the review.

Sources of studies

Relevant studies are sometimes not found, due to poor electronic tagging, and a lack of standardisation of search terms between databases (Magarey 2001). McManus et al. (1998) predicted that only about half of the relevant articles in healthcare are identified by an electronic search; furthermore, many journals are not indexed in MEDLINE (EBBP 2013). In a study describing where papers come from in a systematic review, it was found that only 30% of sources were obtained from the protocol defined at the outset of the study; whereas 'snowballing' identified 51% of the publications and 24% were identified by personal knowledge or personal contacts (Greenhalgh & Peacock 2005). Thus, a singular search strategy focused on electronic databases is insufficient for a comprehensive systematic review. Other potential sources can therefore include:

- Experts in the field (through a defined identification process);
- Conference proceedings (clarifying source meetings and timeframe, as well as the threshold of information needed to allow inclusion);
- References of included studies [manual search of the references cited in the included articles can also reveal other studies missed in the search (Cook & West 2012)];
- Hand searching of key journals in the field.

To avoid selection bias, it is vital that the literature search is comprehensive and whichever sources are selected, they are searched rigorously. This is the most time consuming step of the process. The search strategy for the literature must be carefully documented, and should be carried out with scientific rigour to establish the validity and reproducibility of the review (Magarey 2001; EBBP 2013).

The protocol must address which sources of information will be used. A comprehensive systematic search will usually consider multiple sources of information. Electronic databases should be searched first (Magarey, 2001). For interventional education research, most high-quality primary studies can be identified by four standard databases: the Cochrane Controlled Trial Register, MEDLINE, EMBASE and Social Sciences Citation Indexes (Greenhalgh & Peacock 2005). However, there are hundreds of other databases relevant to specific topics that may also be searched as part of the strategy. Common indexed databases used in medical education include MEDLINE, EMBASE, Scopus, PsycINFO, Web of Science, CINAHL (Cumulative Index to Nursing and Allied Health Literature [for nursing]) and ERIC (Educational Resources Information Centre), to name a few (Cook & West 2012).

Other informal approaches such as browsing, 'asking around' and 'serendipitous discovery' (finding relevant papers when looking for something else) can increase the yield and efficiency of search efforts (Greenhalgh & Peacock 2005), but such sources must be clearly identified prospectively in the protocol and the searching itself recorded so the process is transparent and reproducible.

For a thorough comprehensive search, it is important that foreign language literature is also searched (Magarey 2001; EBBP 2013) Whilst this can be outside the scope of some reviews, omitting articles based on language as opposed to methodology can result in significant bias in the review and should be acknowledged as such. You must also account for publication bias in the search strategy itself. That is, if the review is to be comprehensive, it is important to access various forms of unpublished and grey literature such as theses (Magarey 2001).

Electronic search strategy

This section of the protocol on the search strategy must include what search terms will be used to query the information sources. Knowledge of appropriate indexing terms is required, along with qualifiers and logical operators – these vary in different indexes (Cook & West 2012). Thus, inputs from experts such as research librarians can be immensely helpful. The validity of the preliminary search strategy can be verified by ensuring that known relevant articles are identified using the planned keywords. Reviewers can then look for new keywords in any of the omitted articles to further improve the search strategy. All the articles identified in the search (including those that are excluded at a later stage) must be assigned a unique identification number. The complete search strategy, including specific search terms for all the indexed databases, and other sources should be archived for subsequent reporting. The publication dates covered by the search should also be recorded (Magarey 2001).

When forming the strategy, considering the PICO and CAPS items is once again useful in planning and populating the terms. For example, a systematic review of educational interventions for undergraduate medical students to enhance their skills to handover/handoff between shifts.

(Undergraduate medical OR MBBS OR MB ChB) AND (Course OR Teaching OR Learning OR Intervention) AND (Handover OR Handoff) would structurally form the basis of a search. The amount of terms in each area could then be refined.

Step 7: Performing the search and selecting studies

The inclusion/exclusion process should involve at least two reviewers to minimise random error and bias. This process has two stages.

Stage 1. Reviewers look only at the title, abstract and keywords. Based on these components, if both the reviewers are convinced that the publication is ineligible, the article is excluded. In case there is insufficient information to decide, the article can advance to Stage 2. If reviewers disagree about the inclusion of an article, it is better to take it to Stage 2, based on the rationale that it is better to assess using the full text rather than abstract (Cook & West 2012).

Stage 2. Reviewers read the full text of each article to make final inclusion/exclusion decision. Thus, two independent reviews of all articles are needed to determine which articles will be included. If the reviewers disagree at Stage 2, another

team member may be included to assess the article and consensus must be reached. This is the stage at which concise recording of decision-making is needed to ensure transparency of the review process.

Flow diagrams can be helpful in illustrating this process of searching and choosing primary articles to be included in the review (Crowther & Cook 2007). It is a key to remember at this stage that the quality of an article is not being judged in any way. A study may be of an extremely low quality, but if it meets all the inclusion criteria, it must be included.

Once the literature search is finished and articles have been selected, articles should be reviewed independently by at least two researchers with a detailed documentation of the rejected articles. This is important to disclose the reason for rejection of articles in the completed review (Magarey 2001).

Step 8: Extracting data from the studies

This is the first step in the systematic review process in which there may be some divergence of methodology, depending on the specific type of review being performed. As has already been mentioned, it is outside the scope of these works to offer a full digest on the numerous forms of education systematic reviews available. However, almost all the steps up to this point and many of the remaining are generic and of use to completing systematic review works, regardless of the specific synthesis methods. For the purposes of this section, a standard systematic review methodology will be discussed.

This stage involves three elements that must be extracted in a standardised fashion from all included studies:

(1) *Descriptive data*. Information on the study itself, any educational interventions, curriculum items or assessments used must be recorded.
(2) *Quality assessment*. Key methodological information must be extracted to allow the quality of the study to be assessed.

It is useful to devise a data extraction and appraisal form to facilitate this stage of the review. If completed electronically, this allows for independent extraction of data by a number of authors and then for agreement to be ensured. Once again, if there is disagreement in judgments, a third author should be consulted and consensus reached. Such data extraction forms do exist for use, such as on the BEME website (www.bemecollaboration.org), but it is advised that an individual adaptation is made for each review.

This form should be pre-tested on several studies before commencing the review (Magarey 2001). The contents of the data to be extracted should be defined both conceptually and operationally, with detailed definitions and examples being essential. As the reviewers go through articles during the inclusion process, new questions often emerge, and can also be used to determine which data can be abstracted. As with the inclusion/exclusion process, pilot testing is necessary to identify ambiguous definitions and other areas that may require further clarifications (Cook & West 2012). The data extraction process should be conducted by two reviewers, and disagreements in coding can be resolved through achieving consensus or by including a third reviewer if necessary (Cook & West 2012).

The reviewers must also decide how to account for articles that may have incomplete information. Some possible solutions include: excluding such articles from the review, attaining the missing information from other articles, or trying to obtain the missing information from the original authors themselves. The third of these options is often the most appropriate. Recent work in the clinical domain has demonstrated have often authors do not publish detail regarding the nature of non-pharmacological interventions, but that they were often happy to share such details on request (Hoffmann et al. 2013). Whatever route is selected, the impact of these decisions should be considered in the overall review results (Cook & West 2012).

Descriptive data

This is very much influenced by the focus of the review. The PICO framework can provide guidance on which data to collect (EBBP 2013), including:

- *Key features of participants*: number and key demographics (age range, mean age, sex, race/ethnicity, socioeconomic status);
- *Interventions*: key elements of design, intensity, timing, duration and implementation of intervention;
- *Comparisons*: similar to intervention;
- *Outcomes*: the measurement method and the actual result.

The reviewers should also code information on study design (number of groups, method of randomisation or allocation, allocation concealment method), timing of assessments (post-intervention versus pre- and post-intervention), enrolment and follow up rates, and other features of study quality that can vary for different study designs (Cook & West 2012). It can also be useful to summarise methodological limitations of each study. There are many ways for categorising study quality, including the Medical Education Research Study Quality Instrument (MERSQI) for education research, Jaded Scale for randomized trials, Newcastle-Ottawa Scale for non-randomised studies and Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) for studies of assessment tools (Cook & West 2012). Outcome data recorded may need to be influenced by the pilot and scoping studies and iteratively reconsidered as the work proceeds. The researchers should consider the scope for meta-analysis or qualitative analysis as the data is encountered and attempt to record homogenous data sets.

If performing a review of educational interventions, it is a key to collect information that will allow the nature of the education to be understood and potentially synthesised in analysis:

- Conceptual frameworks or theoretical underpinning for interventions;
- Learning outcomes defined;
- Pedagogy applied;
- Resources and equipment required.

This information is crucial for two main reasons. First, it allows readers to have insight into what education in this area looks like. This is key to support replication and dissemination

and surprisingly is absent from many works in education that are presented in a public arena (Gordon et al. 2013b). Therefore, using a systematic review as a method to collect such data and possibly uncover previously unpublished information regarding such interventions should be seriously considered. Second, such information offers a unique opportunity for the team to consider producing new primary knowledge from this secondary synthesis. So-called Clarification Review Works (Cook et al., 2008) have been successfully used to produce new theoretical understanding and so illuminate the issues at hand and support new innovations from educators (Gordon et al. 2012). Whilst not a pre-requisite for such works or a marker of quality, when considering the issue of impact for educators, such analysis is often a valuable prospect and often available for a relatively modest increase in resource within the wider scope of the review.

Quality assessment of studies

This is an area where opinions in the field do vary. There is general consensus that a single arbiter of quality is rarely relevant in this context and can be misleading (Yardley & Dornan 2012). However, when presenting studies with varied methodology, it is important to consider some key elements of the methodology, to allow readers to judge any potential sources of bias or concern.

Randomised controlled trials are relatively scarce in the field, but given their role within clinical medicine, elements for evaluation are well reported (Evans 1999):

- *Selection bias*: during the research process, concealment and blinding of randomisation are recommended;
- *Performance bias*: any difference in the treatment the subjects receive, other than the intervention being investigated. The treatment of the subjects in both groups should be identical, apart from the intervention that is being evaluated;
- *Attrition bias*: there should not be major differences between the study groups in terms of number of participants that drop out of the study;
- *Detection bias*: the methods used to evaluate the results of the study should be identical for all subjects in the treatment and control group.

When assessing a cohort study, you may wish to consider:

- *Interviewer bias*: knowing exposure status may influence how the outcome is determined;
- *Loss to follow-up bias*: there are similar rates of recruitment, refusal or dropout rates in the two groups;
- *Confounding*: main potential confounders have been identified and accounted for in the design and analysis;
- *Information bias*: the cases and controls have been accurately classified using the same inclusion/exclusion criteria;
- *Recall bias*: it is clearly established that the controls are non-cases.

Many of these issues are applicable for other forms of methodologies, such as before and after studies. However, there are some specific issues that are pertinent to considering quality in medical education reports. Whilst the details for each of these items may have a role in the descriptive components of the analysis, their presence or absence in the report can give an indication of the overall quality of the report or possibly the quality of the reporting.

- Is there a clear research question?
- Are the characteristics of learners or users described?
- Are the descriptive elements of education described above described in a manner that supports replication?
- Are outcomes to be measured described and presented to allow replication?
- How can the outcomes being assessed be characterised (see below)?
- Are the conclusions of the study supported by the methods and results presented (see below)?

The issue of outcomes is often assessed using Kirkpatrick's hierarchy of evaluation (Kirkpatrick & Kirkpatrick 2009). This categorises outcomes in health education at one of several levels – Level 1, satisfaction with education; Level 2a, change in attitudes; Level 2b Change in knowledge; Level 3, change of behaviour in the workplace; and Level 4, change in delivery of care and health outcomes. Because of the pyramidal nature of this framework, it is often understood that higher levels denote higher quality. However, this is not the case, with higher levels more strongly correlated with difficulty of measure, rather than quality. In fact, the question being asked will define the relevance of different levels for the study team. For example, a justification review may be concerned with ensuring a given intervention can change behaviour, where as a clarification review may be more concerned with analysing attitudes or views on the learning experience. Kirkpatrick's hierarchy should not be used as a strict marker of quality, but rather another piece of information that allows a multifaceted assessment of a paper to be made (Yardley & Dornan 2012).

More importantly, the issue of strength of conclusions is an extremely important and yet very subjective item. This item is essentially a judgement as to whether the conclusions of a work are an appropriate reflection of the work completed. Put another way, do the methodology and results fully support the conclusions being made by the researchers. Whilst this may seem self-evident, there is often a mismatch in the conclusions of such works and the conclusions or practice points that are often the items auctioned by readers, making such a judgement is key. As this is essentially a subjective measure, it is important for authors to clarify how such judgements are made. BEME have produced such a measure that may be worth consulting at the planning stage (www.bemecollaboration.org).

Some researchers insist that studies should not be included in systematic review analysis if one or more systematic errors are found (Magarey 2001). Thus, the study findings must be weighted according to their methodological rigour. It is suggested that in the context of education evidence synthesis such exclusions should not be included, given that most studies will be of a method that means such issues are not relevant and indeed, if cohort or controlled studies exist, it is important to consider their contribution. An alternate approach is to include all relevant studies, describe the methodological

issues and possibly perform a subgroup analysis that removes such lower quality studies.

Underlying all of these items of quality must be a key element – heterogeneity. This can be considered in three broad categories.

- Educational or contextual;
- Methodological;
- Statistical.

Educational or contextual heterogeneity describes how different the studies are in regards to their educational environment or context, learner or educator attributes, content, teaching or assessment methods. Methodological heterogeneity is focused on the specific approach of the studies in question. In particular, whether the outcome measures used, time of assessment and basic study design are comparable. Whilst these items are mentioned amongst quality to prompt the extraction of such information, the purpose of considering heterogeneity is to inform the next step. Statistical heterogeneity is discussed in the next section.

A checklist can be developed to summarise these various elements and as such allow readers to evaluate the validity of the studies. Different study designs will require different rating schemes and checklists (EBBP 2013). In addition, it is suggested that a single score or rating is not given to studies. Doing so reduces the complexity of the methodological issues at hand without offering any advantage, other than an imaginary clarity to this innately complex and multifaceted issue. Instead, presenting the judgements made in a single table or graphical representation with perhaps the addition of a traffic light system to indicate areas of concern can be considered, similar to as is used in Cochrane reviews. This offers a subjective, but clear and easily understandable method of presenting such complex data.

The final issue to be considered before moving on from data extraction is how to deal with missing data. In the context of health education and given the extensive nature of the data that must be extracted, it is common to be missing key items. As mentioned previously, it is worth considering contacting authors to allow the data set to be as complete as possible. If attempts are made, regardless of whether successful, the outcomes should be clearly reported.

Step 9: Synthesising and analysing the data

If the previous steps of systematic inclusion and data extraction are conducted appropriately, they facilitate analysing the collected evidence, in whatever form this may take. The important point of this process is to understand that this step involves actually *synthesising* the data, as opposed to cataloguing it (Cook & West 2012). As such, this can been seen as the first step that produces essentially new knowledge.

Synthesis itself goes beyond just reporting the results of each study or counting the number of studies with statistically significant results (Cook & West 2012). Rather, synthesising involves 'pooling and exploring the results to provide a "bottom-line" statement regarding what the evidence supports and what gaps remain in our current understanding' (Cook & West 2012, p. 950). In qualitative data sets, this can involve using similar techniques as are used in primary research to allow the data captured to be clarified and interpreted. With whatever method of synthesis that is used, you must provide transparency in the process that can allow readers to verify the interpretations and reach their own conclusions (Cook & West 2012).

Plan of analysis

A key part of the process is to develop a plan of analysis, created in collaboration with an epidemiologist or a biostatistician, documented in the protocol (Crowther & Cook 2007).

You must make three key decisions regarding the analysis:

(1) Will you statistically pool quantitative results by conducting a meta-analysis? There is no clear decision rule regarding this. It is a judgment call that must be made. If yes, you must decide which statistical model to apply and how to standardise outcomes across studies (Cook & West 2012).
(2) How will heterogeneity or inconsistency across studies be considered and explored? Cook and West (2012) point out that one of the most interesting parts of such reviews is the exploration of why results differ across studies. It is important to explain and report any inconsistencies in between studies (Cook & West 2012) as this can be key in supporting the answering of questions that are deeper than whether interventions are effective, such as when and for whom (Gordon et al. 2013a). In addition, judgements as to the level of heterogeneity of all types must be made to allow the appropriateness or validity of any statistical analysis to be made.
(3) How will you consider threats to the validity of the review? This can be done by transparently reporting the methods, acknowledging key assumptions, exploring potential sources of bias and providing tables containing detailed information on each study assessed. Doing so will help the reader verify and interpret the results themselves. A strong systematic review includes an assessment of the strengths, weaknesses, heterogeneity and gaps in the evidence (Cook & West 2012).

Depending on the aims, questions and broad methodologies of the review, different routes of analysis will be needed.

If it is felt pertinent to complete any meta-analysis, a judgement of the extent of methodological and educational heterogeneity must firstly be made. Whilst there is no objective tool to support this, it is a simple key for the authors to explain their judgement in a transparent manner. If appropriate, the first stage of interpreting such analysis should be the consideration of statistical heterogeneity. Two common measures are the chi squared text and the I2 statistic. These are often automatically produced by the common software packages. If these suggest that there is a high degree of statistical heterogeneity, the authors should consider if the analysis was appropriate. If this is felt to be the case, then consideration of a different statistical test may be needed. For example, a random instead of a fixed effects model.

Qualitative synthesis

If qualitative analysis is to be completed, this needs to be done in the same transparent, stepwise and rigorous manner. Recently, key methodologies that may be applied in this context have been summarised (Bearman & Dawson 2013); these include thematic analysis, meta-ethnography and realist synthesis. Detailed guidance on how to perform such synthesis is outside of the scope of this guide, but it is worth noting that any qualitative methodology can be applied to analyse the evidence that is collated from the review. As long as the analysis is completed within the recognised, transparent and rigorous structure that has been identified and is focussed on answering the aims of the work, all methods are appropriate.

Systematic reviews have often omitted qualitative evidence in favour of quantitative evidence (Dixon-Woods et al. 2005). However, policy makers and practitioners have become increasingly aware of the limitations of using only randomised controlled trials as the single source of valid evidence, and have instead called for more inclusive forms of review (Dixon-Woods et al. 2005). Even the BEME Collaboration (www.bemecollaboration.org) recognises that systematic reviews cannot and should not be limited to randomised controlled trials as they may not be the most appropriate to answer a particular research question (Morrison 2005). Qualitative studies focus on improving understanding of the deeply held views of target groups of an intervention (Thomas et al. 2004). Incorporating qualitative research in systematic reviews continues to present a major challenge. Although the new call for more inclusive evidence has been welcomed, it has also highlighted the limitations of systematic review methodology as being under-developed and under-evaluated (Dixon-Woods et al. 2005). Social scientists continue to address the challenge of synthesising qualitative and quantitative data in a systematic review (Petticrew & Roberts 2005). Integration of the two types of studies can help identify ways that can improve interventions and their implementation (Thomas et al. 2004). Attempts to synthesise qualitative and quantitative evidence generally involve conversion of qualitative data into quantitative form or vice versa. Some of the methods to synthesise qualitative research include the following methods.

Content analysis is a technique for categorising data and determining the frequencies of these categories. It requires that the specifications for the categories be sufficiently precise to allow multiple coders to achieve the same results. It also relies on the systematic application of rules and draws on concepts of validity and reliability. It allows a systematic way of categorising and counting themes, is fairly transparent and easily auditable. Software packages are available for undertaking this analysis. This form of analysis converts qualitative data into quantitative form, making it easier to manipulate within quantitative frameworks. It is often confused with thematic analysis, and is inherently reductive as it tends to diminish complexity and context. It is also unlikely to preserve the interpretive qualities of qualitative evidence. The results may be oversimplified and count only what is easy to classify and count rather than what is actually meaningful and important (Dixon-Woods et al. 2005).

Case survey is a formal process for systematically coding relevant data from a large number of qualitative cases for quantitative analysis. Multiple coders score the cases. One of its main strengths is its ability to synthesise both qualitative and quantitative evidence. Limitations include a reliance on having a sufficient number of cases to make quantitative analysis worthwhile. It also has difficulty in coping with the interpretive properties of qualitative data and is more suited to studies of outcomes rather than processes (Dixon-Woods et al. 2005).

Thematic analysis is one of the key challenges in synthesising qualitative research is translation of concepts between studies (Thomas & Harden 2008). This method involves the identification of prominent or recurrent themes in the literature, and summarising the findings of different studies under thematic headings (Dixon-Woods et al. 2005). It allows for integrating qualitative and quantitative evidence, however it has several limitations: it can either be data driven or theory driven, leading to a possible lack of transparency and there is a general lack of clarity about what exactly this analysis involves and the process by which to achieve it. It is not clear whether it should reflect the frequency with which particular themes are reported, or whether the analysis should be weighted towards themes that appear to have high levels of explanatory value (Dixon-Woods et al. 2005). *Grounded theory* is a primary research approach very influential in development of qualitative methods in health, and it describes the methods of qualitative sampling, data collection and data analysis. It offers an approach for synthesis of primary studies by treating study reports as a form of data on which analysis can be performed using the constant comparative method. As it is concerned with theoretical saturation and theoretical sampling, it also limits the numbers of papers that need to be reviewed as the emphasis is on conceptual robustness rather than completeness of data (Dixon-Woods et al. 2005). Its disadvantage as a method for review is the inherent lack of transparency. It does not offer advice on how to appraise studies for inclusion in a review. And, whilst 'grounded theory' has been used to label many different types of analysis, it should not be underestimated as a means of synthesising primary studies (Dixon-Woods et al. 2005).

Meta-ethnography is a technique developed by Noblit and Hare (1988), it is specifically developed for synthesising qualitative studies. This approach has several advantages including systematic approach combined with the potential for preserving the interpretive properties of primary data (Dixon-Woods et al. 2005). However, it offers no guidance on sampling or appraisal. It is demanding and laborious, and can benefit from development of suitable software (Dixon-Woods et al. 2005). It also runs into the typical problem of transparency.

Realist review. Realist inquiry is based on a realist philosophy of science and considers the interaction between context, mechanism and outcome. From a realist perspective, intervention X is not thought of as having effect size Y with confidence interval Z. Rather, intervention X (e.g. a programme introduced by policymakers who seek to create a particular outcome) alters context (e.g. by making new resources available), which then triggers mechanism(s),

Table 2. Phases in meta-narrative review (Greenhalgh et al. 2005, p. 420).

(1) Planning phase
  (a) Assemble a multidisciplinary research team whose background encompasses the relevant research traditions (an initial scoping phase may be needed before the definitive research team is appointed).
  (b) Outline the initial research question in a broad, open-ended format.
  (c) Agree outputs with funder or client.
  (d) Set a series of regular face-to-face review meetings including planned input from external peers drawn from the intended audience for the review.
(2) Search phase
  (a) Initial search led by intuition, informal networking and 'browsing', with a goal of mapping the diversity of perspectives and approaches.
  (b) Search for seminal conceptual papers in each research tradition by tracking references of references. Evaluate these by the generic criteria of scholarship, comprehensiveness and contribution to subsequent work within the tradition.
  (c) Search for empirical papers by electronic searching key databases, hand searching key journals and 'snowballing' (references of references or electronic citation tracking).
(3) Mapping phase
Identify (separately for each research tradition):
  (a) The key elements of the research paradigm (conceptual, theoretical, methodological and instrumental).
  (b) The key actors and events in the unfolding of the tradition (including main findings and how they came to be discovered).
  (c) The prevailing language and imagery used by scientists to 'tell the story' of their work.
(4) Appraisal phase
Using appropriate critical appraisal techniques:
  (a) Evaluate each primary study for its validity and relevance to the review question.
  (b) Extract and collate the key results, grouping comparable studies together.
(5) Synthesis phase
  (a) Identify all the key dimensions of the problem that have been researched.
  (b) Taking each dimension in turn, give a narrative account of the contribution (if any) made to it by each separate research tradition;
  (c) Treat conflicting findings as higher-order data and explain in terms of contestation between the different paradigms from which the data were generated.
(6) Recommendations phase
Through reflection, multidisciplinary dialogue and consultation with the intended users of the review:
  (a) Summarise the overall messages from the research literature along with other relevant evidence (budget, policymaking priorities, competing or aligning initiatives)
  (b) Distil and discuss recommendations for practice, policy and further research.

which produce both intended and unintended outcomes. Intervention X may work well in one context but poorly or not at all in another context. A realist synthesis (or realist review – these terms are synonymous) applies realist philosophy to the synthesis of findings from primary studies that have a bearing on a single research question or set of questions (Wong et al. 2013).

Meta-narrative approach. When we are dealing with large multidisciplinary bodies of literature, one of the difficulties that can quickly arise is the sorting of primary studies into a single theoretical taxonomy – theoretical basis might be different, and there might be no unifying principle or a consistent approach to research design (Greenhalgh et al. 2005). This is because each of the bodies of literature may have different paradigmatic lenses in four dimensions.

Conceptual: what is considered to be the important objects of study and what counts as a legitimate problem to be solved;
Theoretical: how the objects of study are considered to relate with one another and to the world;
Methodological: accepted ways in which the problem can be investigated;
Instrumental: the accepted tools and techniques used to investigate the problem.

Table 2 provides a guideline of the different phases in conducting a meta-narrative review.

Subgroup and sensitivity analysis

Rigour of a systematic review can be increased through sensitivity analysis – measuring the impact of the results after adjustment of one or more characteristics of the studies. The strength of inference is much greater if the results are unchanged under varying conditions (Crowther & Cook 2007). Examples of sensitivity analyses include: comparing the pooled results of the lower versus higher methodologically rigorous studies; and measuring and comparing the results using different techniques to impute missing data (Crowther & Cook 2007).

Similarly, subgroup analysis can increase the relevance of results by further scrutinising the data. Such analysis allows the impact of learners, environment, teaching methods or any other factors on the results to be considered. This can be very useful since answering the questions such as when, for who or where can be more informative than focusing on whether interventions are effective.

Step 10: Discussing and concluding the review

Giving specific guidance in this area is difficult, as it will be very much a construct of earlier sections. However, the key elements that must be addressed in this area are:

- Present the main findings in a manner that is related to the questions;
- Discuss strengths and limitations of the review and its findings, commenting on the strength of the evidence base;
- Discuss the implications of the findings for educators and researchers.

Whilst discussing the findings of the review, it is a key to relate this to the review objectives; this might often be a

superfluous item for authors. Commenting on the limitations should particular relate how the quality of primary extracted data has impacted and possibly limited the strength of conclusions made.

Conclusions should be in terms of what the implications are for both educators and researchers. The review team have an extremely in-depth knowledge of the state of the field and as such are very well placed to highlight explicitly any directions for future work. Insights should not be limited to the clinical teacher, but where appropriate, give suggestions for curriculum developers and educational policy makers. This allows reviews to be relevant at both the micro- and macro-educational level. In this way, the authors should essentially seek to identify how the results can be translated into practice. This section therefore becomes one of the least objective and paradoxically most important elements of the reported work. All too frequently this element of the systematic review is missing the transferability of the findings, how they can be used for and inform future educational practice.

Step 11: How the review will be reported

A systematic review report is based on the search for studies that address a clearly defined question, a critical appraisal of the studies and the synthesis of the research findings (Moher et al. 2007); a well-executed review can be of limited value if poorly reported.

Key elements of reporting systematic reviews and meta-analyses can be found in several guidelines such as the preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement (Liberati et al. 2009). Whilst many of these items are key for reporting all systematic reviews, recent work has identified that much of the health education systematic review that essentially follows such reporting guidance is still of limited value to readers (Gordon 2014). As such, specific guidance in this field was produced, the **ST**ructured appr**O**ach to the **R**eporting **I**n healthcare education of **E**vidence **S**ynthesis (STORIES) statement (Gordon & Gibbs 2014; Figure 1). As well as supporting reporting by

The STORIES statement

STORIES statement: Publication standards for healthcare education evidence synthesis

Title and abstract
Use a title that includes a description of the aims of the piece (educational effectiveness, descriptive, etc) and method of evidence synthesis (e.g. realist, meta-ethnographic, etc)
Provide a structured summary

Introduction:
Describe the rationale for the review in the context of what is already known
Provide a statement of the questions being addressed by the study
State why this method of evidence synthesis was selected within the context of the questions being asked

Methodology:
State and provide a rationale for how the searching was done
Provide details on all the sources of information and dates searched
Electronic databases - provide full search terms for at least one database, with details of deviations in subsequent searches
Describe the process of data extraction and any process of contacting authors for confirmation of / or more data
Explain the method for judging inclusion / exclusion
If quality appraisal tools are used, please describe and justify their choice
Describe qualitative methods for synthesising primary evidence (where appropriate) and the goal of these methods, such as thematic analysis; meta-ethnography, and realist synthesis
Describe quantitative methods for synthesising primary evidence (where appropriate), such as meta-analysis and how issues of heterogeneity will be considered

Results:
Give a flow diagram summarising study selection
If individuals familiar with the relevant literature and/or topic area were contacted, provide a summary of the contact and information obtained
Provide summarised details of included works, considering elements such as methodology, key results and conclusions
Describe methods of quality assessment of education reported, including all parameters considered (e.g. Details of study theoretical underpinning, pedagogical strategies and details of teaching activities to allow replication or dissemination)
Describe quality assessment of the research methods of included studies
Present the results of qualitative and/or quantitative evidence synthesis

Discussion:
Present the main findings in light of the review objectives
Discuss strengths and limitations of the review and its findings, commenting on the strength of the evidence
Discuss how the findings of the evidence synthesis impact future primary research
Describe possible implications of the findings for educators

Morris Gordon and Trevor Gibbs 2014, and at http://clok.uclan.ac.uk/9973/

Figure 1. The STORIES statement.

authors, the STORIES report can be used to support critical appraisal of such reports.

The key questions to consider when evaluating systematic reviews are as follows (Bigby & Williams 2003; Crowther & Cook 2007).

Are the results valid?

A specific and focused question that specifies the characteristics of the participants, the nature of the intervention to which the participants will be exposed to and the outcomes that will be measured.

The methods of literature review are explicitly presented with enough clarity and transparency for the reader to determine if important, relevant studies have been omitted from the analysis.

An explicit inclusion/exclusion criteria.

The reader is able to determine from the abstracted data whether the studies assessed were methodologically valid.

The systematic review includes possible reasons for differences among study results.

Are the valid results important?

The degree to which limitations in the analysis are identified and addressed by the author.

The overall results of the review in term of magnitude of benefit or harms.

What are the limitations of a systematic review?

Poor reporting quality

Recent studies assessing the qualities of systematic reviews have found that the quality of reporting is less than optimal (Moher et al. 2007). After capturing a cross-sectional sample of all recently published systematic reviews, Moher et al. (2007) evaluated the reviews in terms of epidemiological and reporting characteristics. They found some disappointing results such as the lack of assessment of publication bias, despite the evidence for its existence and its ability to influence the results of the reviews. More importantly, they found little improvement in the quality of reporting for non-Cochrane reviews, with many not reporting key aspects of the systematic review methodology, diminishing confidence in their results and conclusions (Moher et al. 2007). Moher et al. (2007) comment it is possible that poor reporting of categories maybe reflective of the inadequate guidance available to authors to do systematic reviews. For example, they found that a third of the systematic reviews they examined did not report on how the quality of the studies found in their search had been assessed. Assessment of 'publication bias' was only reported in a quarter of the systematic reviews reviewed. Thus, Moher et al. (2007) conclude that due to the lack of standardised reporting of systematic reviews, readers should not accept the conclusions of systematic reviews uncritically. As already stated, the specific needs in health education systematic review require specific reporting guidance and the recent production of the STORIES statement should be helpful to authors (Gordon 2014).

Outdated systematic reviews

An inherent limitation of a systematic review is that the utility of the reviews diminish over time as the literature becomes outdated (Moher et al. 2007). Due to this limitation, Moher et al. (2007) reflect that journals may hesitate in publishing updates that are substantially the same as previous publications. However, if systematic reviews are to maintain their usefulness, updating them needs to be a high priority (Moher et al. 2007). French et al. (2005) remind us that failure to update reviews can lead to decision makers acting on out of date information. However, on the other hand, reviews that are updated too soon may be a waste of effort and resources, or introduce bias. For example, systematic reviews with few studies are susceptible to the 'time lag bias' – when trials with positive results are published more quickly than those with null or negative results. Another danger of updating too frequently is that repeated significance tests can lead to inflated Type I error (French et al. 2005). In their study of assessing how conclusions can change when Cochrane systematic reviews are updated, French et al. (2005) found that in many cases, updating the reviews did not result in changing the conclusion, nor did it lead to a more precise conclusion. They further suggested that rather than a time based approach to update a review, a priority-setting approach may be more appropriate (French et al. 2005).

Limited datasets and the strengths of conclusions based on a systematic review

If the literature base is very small to begin with, the primary studies may be underpowered and subsequently, so will the conclusions of a systematic review. This is because the review is not independent of the quality of the contributing studies, and if the primary studies are limited or poor, the review will also be flawed (Crowther & Cook 2007).

Inclusion of unpublished data

One of the contentious topics in conducting a systematic review is the inclusion of unpublished data (Roberts & Schierhout 1997). For example, it has been argued that routine inclusion of unpublished data can expose the systematic review to data of lesser quality, since it has not been peer reviewed, and it will be difficult to identify all the sources of the data. It may also be generated using less rigorous techniques, making it more prone to bias (Crowther & Cook 2007). Conversely, others argue that due to publication bias, studies that do not show statistically significant differences or not favouring the drug intervention/activity) (or perhaps educational intervention/activity) under investigation, tend to not get published (Trespidi et al. 2011). Thus, it is important to consider and discuss the inclusion of unpublished data with the team members and a decision should be recorded in the protocol. The reviewers can compare the protocol with the unpublished results to determine if they have deviated from their original analysis plan or not (Crowther & Cook 2007).

Influence of external agencies

Although perhaps not always applicable to educational practice, many systematic reviews are funded by organisations such as pharmaceutical companies (in clinical practice) or special interest groups. By manipulating the inclusion/exclusion study criteria, the design of the systematic reviews can be influenced to select only specific set of studies, biasing the review. Furthermore, the results themselves can be interpreted through the biased lens of reviewers who are influenced by a particular industry (Crowther & Cook 2007). Although without any sound or evidential basis, it is possible that with the rapid growth of the 'simulation industry', similar bias could be introduced into educational research.

Language of publication

Although much medical research is published in English, English speaking reviewers can be restricted in their access to articles in other languages that may be important in their field of interest (Lang 2004).

Limited funding

Due to the ambiguity around the consideration of systematic reviews as original research, limited funding opportunities available to conduct a systematic review project remains a barrier (Meerpohl et al. 2012). A formal agreement on the status of systematic reviews can motivate researchers to undertake such projects on a larger extent. If they are recognised as original research, funding agencies may also begin to provide more financial support (Meerpohl et al. 2012).

Long duration

Finally, systematic reviews can take up to or over 12 months to complete, due to the rigorous process of data collection and peer review. Thus, they are better suited for being part of a larger research study with longer time frames (White & Waddington 2012).

Updating a systematic review

As stated above, this is often a significant limitation in the utilisation of a systematic review. Even though these reviews are often advocated as the best source of evidence available to educators and decision-makers, they are often required to be frequently updated and within a relatively short time (Shojannia et al. 2007). Whilst the Cochrane Collaboration updates its systematic reviews routinely, this appears not to be the case with 80% of all published reviews (Moher et al. 2006). Corrections or re-analysis of a previously conducted systematic review without search for new evidence cannot be considered an update (Moher et al. 2007). Instead, extending a search to new sources, or an exhaustive but fruitless search for new evidence can be still considered an update (Moher et al. 2007). Whilst updating a systematic review can provide important information, this process can nonetheless be as costly and time consuming as conducting the original review. Furthermore, research priorities can also change over time, teams frequently become disbanded and re-form, whilst funding priorities for relevant stakeholders may change (Nasser & Welch 2013).

Conclusions

The use of evidence synthesis and systematic review within medical education has been growing exponentially. In this guide, we have sought to offer a practical digest on the key issues and challenges involved in such an endeavour. Paramount throughout the entire process is clarity of the review question and ensuring this question is relevant to those in the field, in this case, medical education.

Given the nature of medical education, a shift from focussing on questions considering effectiveness of education is needed. Works investigating constituents of education within the evidence so as to allow rapid replication of quality works and delving into deeper clarification questions to offer insight at a rich conceptual level are needed more.

We believe this guide will support the reader in completing such works and delivering reviews that can impact both educators and policy makers alike, in the quest to develop quality evidence-based medical education.

Notes on contributors

RICHA SHARMA, BA, MA, is a PhD student at the School of Population and Public Health, in the Faculty of Medicine at the University of British Columbia. She has worked on child, adolescent and maternal health projects in urban and rural India, and the limitations of law to address gender based violence in India.

MORRIS GORDON, MB.ChB, M.Med, PhD, FRCPCH, FHEA, FAcadMed, is a consultant Paediatrician in Blackpool Victoria Hospital, UK and the Head of Medical Education Research and Transition to Clinical Practice in the Medical School, University of Central Lancashire, UK. His research interests are education in non-technical skills and evidence synthesis in healthcare. He is an active member of the Best Evidence in Medical Education (BEME) Editorial Board.

SHAFIK DHARAMSI, BEd, MSc, PhD, is an Associate Professor with the Department of Family Practice, Research Lead for the Family Medicine Residency Programme at the St. Paul's Hospital Site, and Lead Faculty of the Social Accountability and Community Engagement initiative in the Faculty of Medicine at the University of British Columbia. His scholarship focuses on the social accountability of medical schools, medical education and global health.

TREVOR GIBBS, MD, DA, FRCGP, MMedSc, SFHEA, is an independent consultant and Professor in Medical Education and Primary Care and Development Officer for the Association for Medical Education in Europe (AMEE). He has extensive international experience in curriculum development and community-based education and is an active member of the Best Evidence in Medical Education (BEME) Board.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

References

Altman DG. 1994. The scandal of poor medical research. BMJ 308:283–284.

Bath PMW, Gray LJ. 2009. Systematic reviews as a tool for planning and interpreting trials. Int J Stroke 4(1):23–27.

Bearman M, Dawson P. 2013. Qualitative synthesis and systematic review in health professions education. Med Educ 47:252–260.

Best Evidence Medical Education (BEME). 2014. [Accessed September 2014] Available from www.bemecollaboration.org.

Bigby M, Williams H. 2003. Appraising systematic reviews and meta-analyses. Arch Dermatol 139(6):795–798.

Bligh DA. 2000. What's the use of lectures? San Francisco: Jossey-Bass.
Burgers JS, Grol R, Klazinga NS, Makela M, Zaat J. 2003. Towards evidence-based clinical practice: An international survey of 18 clinical guideline programs. Int J Qual Healthc 15(1):31–45.
Campbell Collaboration. 2013. What is a systematic review? [Accessed 11 November 2013] Available from http://www.campbellcollaboration.org/what_is_a_systematic_review/index.php.
Carline JD. 2004. Funding medical education research: Opportunities and issues. Acad Med 79(10):918–924.
Chalmers I, Glasziou P. 2009. Avoidable waste in the production and reporting of research evidence. Lancet 374:86–89.
Chen FM, Burstin H, Huntington J. 2005. The importance of clinical outcomes in medical education research. Med Educ 39(4):350–351.
Cherry MG, Fletcher I, O'Sullivan H, Shaw N. 2012. What impact do structured educational sessions to increase emotional intelligence have on medical students? BEME Guide No. 17. Med Teach 34(1):11–19.
Clarke M, Hopewell S, Chalmers I. 2010. Clinical trials should begin and end with systematic reviews of relevant evidence: 12 Years and waiting. Lancet 376(9734):20–21.
Cochrane Collaboration. 2013. Cochrane reviews. [Accessed 11 November 2013] Available from http://www.cochrane.org/cochrane-reviews.
Cook DA, Bordage G, Schmidt H. 2008. Description, justification, and clarification: A framework for classifying the purposes of research in medical education. Med Educ 42:128–133.
Cook DA, West CP. 2012. Conducting systematic reviews in medical education: A stepwise approach. Med Educ 46(10):943–952.
Cresswell J, Klassen AC, Clark VLP, Smith KC. 2010. Best practices for mixed methods research in the health sciences. [Accessed 14th October 2013] Available from http://obssr.od.nih.gov/mixed_methods_research/pdf/Best_Practices_for_Mixed_Methods_Research.pdf.
Crowley P, Chalmers I, Keirse MJ. 1990. The effects of corticosteroid administration before pre-term delivery: An overview of the evidence from controlled trials. Br J Obstet Gynaecol 97:11–25.
Crowther MA, Cook DJ. 2007. Trials and tribulations of systematic reviews and meta-analyses. Hematology Am Soc Hematol Educ Prog 2007: 493–497.
Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A. 2005. Synthesising qualitative and quantitative evidence: A review of possible methods. J Health Serv Res Policy 10(1):45–53.
Doshi P, Jones S, Jefferson T. 2012. Rethinking credible evidence synthesis. BMJ 344:d7898.
Drabkin I. 1957. Medical education in ancient Greece and Rome. J Med Educ 32(4):286–296.
Evans D. 1999. Conducting a systematic review. Adelaide, Australia: The Joanna Briggs Institute for Evidence Based Nursing and Midwifery. pp 1–40.
Evidence-Based Behavioral Practice. 2013. Systematic review module: Introduction to systematic reviews. [Accessed 11 November 2013] Available from http://www.ebbp.org/training.html.
French SD, McDonald S, McKenzie JE, Green SE. 2005. Investing in updating: How do conclusions change when Cochrane systematic reviews are updated? BMC Med Res Methodol 5(1):33.
Gordon M. 2014. Developing healthcare non-technical skills training through educational innovation and synthesis of education research. [Accessed 1 October 2014] Available from http://usir.salford.ac.uk/30826/1/M_GORDON_THESIS_FINAL_AMMENDED_FULL_SUBMISSION.pdf.
Gordon M, Darbyshire D, Baker P. 2012. Educational interventions to improve patient safety: A systematic review. Med Educ 46: 1042–1054.
Gordon M, Darbyshire D, Baker P. 2013a. Separating the wheat from the Chaff. Systematic review in medical education. Med Educ 47(6):632.
Gordon M, Darbyshire D, Saifuddin A, Vimalesvaran K. 2013b. Limitations of poster presentations reporting educational innovations at a major international medical education conference. Med Educ Online 18. doi:10.3402/meo.v18i0.20498.
Gordon M, Gibbs T. 2014. STORIES statement Publication standards for health care education evidence synthesis. BMC Med 12:143.
Gordon M, Vaz Carneiro A, Patricio M, Gibbs T. 2014. Missed opportunities in health care education evidence synthesis. Med Educ 48:644–645.
Greenhalgh T, Peacock R. 2005. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: Audit of primary sources. BMJ (Clin Res Ed.) 331(7524):1064–1065.
Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O, Peacock R. 2005. Storylines of research in diffusion of innovation: A meta-narrative approach to systematic review. Soc Sci Med 61(2):417–430.
Gutiérrez Castrellón P, Polanco Allué I, Salazar Lindo E. 2010. An evidence based Iberic-Latin guideline for acute gastroenteritis management in infants and prescholars [Article in Spanish]. An Pediatr (Barc) 72(3):220.e1–220.e20.
Haig A, Dozier M. 2003a. BEME Guide No 3: Systematic searching for evidence in medical education – Part 1: Sources of information. Med Teach 25:352–363.
Haig A, Dozier M. 2003b. BEME Guide No 3: Systematic searching for evidence in medical education – Part 2: Constructing searches. Med Teach 25:463–484.
Hall JC, Platell C. 1997. Half life of truth in surgical literature. Lancet 350(9093):1752.
Hammick M, Dornan T, Steinhart Y. 2010. Conducting a best evidence systematic review. Part 1: From idea to data coding. BEME Guide No. 13. Med Teach 32(1):3–15.
Harden R, Grant J, Buckley G, Hart IR. 1999. Best evidence medical education. BEME Guide No. 1. Med Teach 21(6):553–562.
Hoffmann TC, Erueti C, Glasziou PP. 2013. Poor description of non-pharmacological interventions: Analysis of consecutive sample of randomised trials. BMJ 347:f3755.
Khan KS, Kunz R, Kleijnen J, Antes G. 2003. Five steps to conducting a systematic review. J R Soc Med 96(3):118–121.
Kirkpatrick JD, Kirkpatrick WK. 2009. Then and now: A strong foundation for the future. St Louis, MO: Kirkpatrick Partners.
Lang TA. 2004. The value of systematic reviews as research activities in medical education. Acad Med 79(11):1067–1072.
Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. BMJ 339:B2700.
Magarey JM. 2001. Elements of a systematic review. Int J Nurs Pract 7(6):376–382.
Mbuagbaw L, Wiysonge CS, Nsagha DS, Ongolo-Zogo P, Pantoja T. 2011. An introduction to systematic reviews and meta-analysis: A workshop report on promoting evidence based medical practice through capacity building in research synthesis. Pan Afr Med J 8(1):15.
McManus R, Wilson S, Delaney B. 1998. Review of the usefulness of contacting other experts when conducting a literature search for systematic reviews. BMJ 317:1562–1563.
Meerpohl JJ, Herrle F, Antes G, Von Elm E. 2012. Scientific value of systematic reviews: Survey of editors of core clinical journals. PLoS One 7(5):e35732.
Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, Pham B, Klassen TP. 1999. Assessing the quality of reports of randomised trials: Implications for the conduct of meta-analyses. Health Technol Assess 3(12):i–iv, 1–98.
Moher D, Tetzlaff J, Tricco A, Sampson M, Altman D. 2006. Epidemiology and reporting characteristics of systematic reviews. PLoS Med 4:e78.
Moher D, Tsertsvadze A, Tricco A, Eccles M, Grimshaw J, Sampson M, Barrowman N. 2007. A systematic review identified few methods and strategies describing when and how to update systematic reviews. J Clin Epidemiol 60(11):1095–1104.
Morrison J. 2005. The challenges of systematic reviews of educational research. BMJ (Clin Res Ed) 331(7513):391.
Mulrow CD. 1987. The medical review article: State of the science. Ann Intern Med 106(3):485–488.
Nasser M, Welch V. 2013. Prioritization of systematic reviews leads prioritization of research gaps and needs. J Clin Epidemiol 66(5):522–523.
Noblit GW, Hare RD. 1988. Meta-ethnography: Synthesizing qualitative studies. London: Sage.

Norman G. 2003. RCT = results confounded and trivial: The perils of grand educational experiments. Med Educ 37:582–584.

Olsen O, Middleton P, Ezzo J, Gøtzsche PC, Hadhazy V, Herxheimer A, Kleijnen J, McIntosh H. 2001. Quality of Cochrane reviews: Assessment of sample from 1998. BMJ 323:829–832.

Patricio M, Vaz Careneiro A. 2012. Systematic reviews of evidence in medical education and clinical medicine: Is the nature of evidence similar? Med Teach 34:474–482.

Petticrew M, Roberts H. 2005. Systematic reviews in the social sciences: A practical guide. London: Wiley-Blackwell.

Pitkin RM, Branagan MA, Burmeister LF. 1999. Accuracy of data in abstracts of published research articles. J Am Med Assoc 281(12):1110–1111.

Poynard T, Munteanu M, Ratziu V, Benhamou Y, Di Martino V, Taieb J, Opolon P. 2002. Truth survival in clinical research: An evidence-based requiem? Ann Intern Med 136(12):888–895.

Roberts I, Schierhout G. 1997. The private life of systematic reviews. BMJ 315:686–687.

Sackett DL, Rosenberg WM. 1995. On the need for evidence-based medicine. J Public Health 17(3):330–334.

Sackett DL, Rosenberg WM, Gray JAM, Richardson WS. 1996. Evidence based medicine: What it is and what it isn't. BMJ 312:71–72.

Shactman N. 2000. Why doctors hate the Internet. Wired News. [Accessed 14 October 2013] Available from http://www.wired.com/news/print/0,1294,35516.00.html.

Shojannia KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. 2007. How quickly do systematic reviews go out of date? A survival analysis. Ann Intern Med 147(4):224–233.

Smith V, Devane D, Begley CM, Clarke M. 2011. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC Med Res Methodol 11(1):15.

Social Science Research Unit. 2009. About EPPI centre. [Accessed 14 October 2013] Available from http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=63.

Thomas J, Harden A, Oakley A, Oliver S, Sutcliffe K, Rees R, Kavanagh J. 2004. Integrating qualitative research with trials in systematic reviews. BMJ 328(7446):1010–1012.

Thomas J, Harden A. 2008. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 8(1):45.

Todres M, Stephenson A, Jones R. 2007. Medical education research remains the poor relation. BMJ 335(7615):333–335.

Trespidi C, Barbui C, Cipriani A. 2011. Why it is important to include unpublished data in systematic reviews. Epidemiol Psychiatr Sci 20(2):133–135.

White H, Waddington H. 2012. Why do we care about evidence synthesis? An introduction to the special issue on systematic reviews. J Dev Eff 4(3):351–358.

Wong G, Greenhalgh T, Westhopr G, Buckingham J, Pawson R. 2013. RAMESES publication standards: Realist syntheses. BMC Med 11:21.

Yardley S, Dornan T. 2012. Kirkpatrick's levels and education 'evidence'. Med Educ 46:97–106.

Yousefi-Nooraie R. 2009. Promoting development and use of systematic reviews in a developing country. J Eval Clin Pract 15(6):1029–1034.