Real world reviews: a beginner’s guide to undertaking systematic reviews of public health policy interventions

C Bambra

ABSTRACT

Background The systematic review is becoming an increasingly popular and established research method in public health. Obtaining systematic review skills are therefore becoming a common requirement for most public health researchers and practitioners. However, most researchers still remain apprehensive about conducting their first systematic review. This is often because an ‘ideal’ type of systematic review is promoted in the methods literature.

Methods This brief guide is intended to help dispel these concerns by providing an accessible overview of a ‘real’ approach to conducting systematic reviews. The guide draws upon an extensive practical experience of conducting various types of systematic reviews of complex social interventions.

Results The paper discusses what a systematic review is and how definitions vary. It describes the stages of a review in simple terms. It then draws on case study reviews to reflect on five key practical aspects of the conduct of the method, outlining debates and potential ways to make the method shorter and smarter—enhancing the speed of production of systematic reviews and reducing labour intensity while still maintaining high methodological standards.

Conclusion There are clear advantages in conducting the high quality pragmatic reviews that this guide has described: (1) time and labour resources are saved; (2) it enables reviewers to inform or respond to developments in policy and practice in a timelier manner; and (3) it encourages researchers to conduct systematic reviews before embarking on primary research. Well-conducted systematic reviews remain a valuable part of the public health methodological tool box.

The systematic review, which has been advanced for many years now in the evidence-based medicine literature, is becoming an increasingly popular and established research method in public health as well as some of the social sciences. For example, a quick search of the Web of Science database obtains 707 hits for the subject ‘systematic review’ in the social science literature for the period 1945—99 compared with over 4596 hits for 2000—8. This includes an increase from 85 hits to 764 hits for the subcategory public, environmental and occupational health. In part, this rapid expansion in the use of systematic reviews can be mainly attributed to the increased emphasis in public health policy and practice for decisions and interventions to be more ‘evidence based’, the so-called ‘ascendancy of evidence’. The systematic review is now widely considered within policy and practice circles to be a good way of making the sometimes conflicting and complicated results of many different types of study accessible and more useable. For example, a recent UK government commissioned review of promoting public health highlighted the value of systematic reviews in providing robust and reliable evidence on the effectiveness of interventions. Similarly, they are also a key factor in the formulation of binding recommendations made for the UK NHS by the National Institute for Health and Clinical Excellence.

Obtaining, often on the job, skills in systematic review methodology (both conducting them and interpreting them) is therefore becoming a common requirement for most public health researchers and practitioners. This is reflected in the number of guidebooks available. However, most researchers still remain apprehensive about conducting their first systematic review and the voluminous guidebooks and handbooks (most are hundreds of pages long) do little to quell such feelings. The fact that most of the commentaries on conducting systematic reviews still come from the healthcare or evidence based medicine literatures also makes the method inaccessible to public health policy researchers. Systematic reviews of public health interventions draw on an evidence base dominated by observational and qualitative studies in which the measurement of relevant outcomes is often heterogeneous. This means that the advice on systematic reviews from the healthcare and medicine literature, concerned as it is with the quality of experimental (particularly randomised controlled trial) studies and the meta-analysis of results, is difficult for novices to transfer to public health questions. Looking to social science is not particularly helpful either as a lot of the systematic review literature in this field still highlights the difficulties, as opposed to the advantages, of applying the method in social evaluation. From the outside looking in, the systematic review approach can initially seem time consuming, tedious and overwhelming. Added to this, there are various debates in the systematic review literature about what a systematic review is, how it should be conducted, and by whom.

This brief guide is intended to help dispel these concerns by providing an accessible, more tailored overview for novices of the systematic review literature.
WHAT IS A SYSTEMATIC REVIEW?
The first issue that any researcher is faced with is establishing what counts as a systematic review. In simple terms the systematic review is a method of locating, appraising and synthesising evidence. However, in practice there are disagreements about what does and what does not constitute a systematic review. Conventional public health and social science definitions of the systematic review tended to concentrate on distinguishing it from traditional (non-systematic) literature reviews. So, for example, Oakley and Fullerton23

Box 1 Case study systematic reviews

Completed systematic reviews

**Effectiveness of the welfare to work programmes for people with a chronic illness or disability (2005)\(^\text{14}\)**
Quantitative and qualitative studies of the employment effects of UK welfare to work programmes directed at people with a disability or a chronic illness were identified using 17 electronic databases, hand searches of the relevant literature, searches of the world wide web, citation follow-up and contacts with authors. Sixteen qualitative and quantitative observational studies were included and critically appraised.

**Interventions that aim to increase employee participation or control (2007)\(^\text{15}\)**
Experimental and quasi-experimental studies reporting health and psychosocial effects of interventions that increased employee participation or workplace control were examined. Seventeen electronic databases (medical, social science and economic), bibliographies and expert contacts were searched. Eighteen observational studies were included and critically appraised.

**Workplace task restructuring interventions (2007)\(^\text{16}\)**
Experimental and quasi-experimental studies reporting health and psychosocial effects of changes to the work environment brought about by task structure work reorganisation were examined. Seventeen electronic databases (medical, social science and economic), bibliographies and expert contacts were searched. Nineteen observational studies were reviewed.

**The health effects of reorganising shift work (2008)\(^\text{17}\)**
Systematic review of experimental and quasi-experimental studies that evaluated the effects on health and work—life balance of organisational-level interventions that redesign shift work schedules. Twenty-seven electronic databases (medical, social science, economic) were searched. Twenty-six observational studies were synthesised.

**The effects of compressed work week interventions on the health and wellbeing of shift workers (2008)\(^\text{18}\)**
Studies of the effects of the compressed working week on the health and work—life balance of shift workers were systematically reviewed. Twenty-seven electronic databases were searched as well as websites, bibliographies and expert contacts. Forty observational studies were included.

**The health effects of volunteering (2008)\(^\text{19}\)**
Systematic review of the health effects of volunteering on individual volunteers and on health service users. Eleven electronic databases were searched for qualitative and quantitative studies (both intervention studies and comparative studies). Expert contacts were used. Eighty-seven qualitative and observational quantitative papers were included in the review.

**The effects on health and health inequalities of partnership working (2009)\(^\text{20}\)**
Systematic review of quantitative (longitudinal before and after) and qualitative studies (1997—2008) reporting on the health (and health inequalities) effects of public health partnerships in England. Eighteen electronic databases (medical, social science and economic) were searched as well as websites, bibliographies and expert contacts. Fifteen studies were reviewed.

Ongoing systematic reviews

**Return-to-work interventions for people with a chronic illness**
Update of 2005 review of the effectiveness of welfare to work interventions.14 Sixteen electronic databases have been searched and 40 studies are currently being synthesised.

**Psychosocial work environment and lower back pain**
Systematic review of prospective cohort studies of the association between the psychosocial work environment and the development of lower back pain. Four electronic databases as well as bibliographies have been searched.

**Flexible working conditions and their effects on employee health and wellbeing**
A Cochrane collaboration systematic review of the effects of flexible working conditions on the health of workers. Six electronic databases have been searched as well as bibliographies, websites and expert contacts.

J Epidemiol Community Health 2011;65:14—19. doi:10.1136/jech.2009.088740

---

method and its place in public health policy research. The guide draws upon an extensive practical experience of conducting various types of systematic reviews of complex social interventions in the field of public health policy (box 1) and in terms of training researchers.14—22 First, it discusses what a systematic review is and describes the stages of a review in simple terms. It then reflects on five key practical aspects of the conduct of the method, outlining debates and potential ways to make the systematic review method shorter and smarter. It concludes by highlighting the ongoing value of the method.
defined the systematic review as ‘a review of a clearly formulated question that uses systematic and explicit methods to identify, select and critically appraise relevant research, and to collect and analyse data from the studies that are included in the review’. Similarly, a systematic review is systematic because it attempts systematically to locate research, both published and unpublished, and critically evaluate it on grounds of relevancy and predetermined methodological criteria. Only research that is judged to be both relevant to the review question, and that fulfils the methodological inclusion criteria, is combined into the final review analysis. The systematic review combines the results of these studies and thus provides a summary of the ‘best available evidence’ on a given question.11 12 Following the database of abstracts of reviews of effects guidelines on the minimum requirements of a systematic review,24 recent umbrella reviews (systematic reviews of reviews) have tended to define studies as systematic reviews if they addressed a clearly defined question, and an effort had been made to identify all relevant literature by searching at least one named database combined with either checking references, hand searching, citation searching, or contacting authors in the field.25-27 Although this is clearly only a minimal definition of a systematic review and not a definition of good practice, it does illustrate that systematic reviews do not therefore need to search everywhere and synthesise all studies on a given question. In these diverse definitions, the clear distinguishing features of the systematic review are therefore its formality, transparency and replicability: its ‘systematic-ness’.

STAGES OF A SYSTEMATIC REVIEW

A good quality systematic review follows a formal procedure that starts with the formulation of a precise question that includes a definition of the participants, the intervention to be assessed and the outcomes to be measured.11 12 This is followed by the development of a protocol that a priori outlines the inclusion and exclusion criteria, the searches to be undertaken and the sources to be consulted. A rigorous literature search, which combines electronic database searches, searching the references of identified studies, hand searching relevant journals and contacting experts in the field, is then conducted.11 12 Studies are then selected in terms of whether they meet the strict inclusion criteria (which usually includes criteria to define relevance to the review question and methodological criteria).

Included studies are then data extracted using standardised forms and subjected to a methodological critical appraisal (in which the quality of the studies is assessed). The studies are then synthesised using meta-analysis (a statistical strategy for pooling the results of several studies into a single effect estimate) or narrative synthesis (data synthesis and exploration of heterogeneity using description) taking into account study designs and the results of the critical appraisal. The results of the synthesis then need to be interpreted and reflected upon in terms of the research, policy and practice implications. The final stage of the review is writing up and publication, and it is essential that the methods of the systematic review are clearly described and the limitations of the included studies (and the systematic review itself) are acknowledged.

CONDUCTING A REVIEW

These suggestions on how to conduct a systematic review are based on experience of 10 reviews (box 1) and are related to different stages and aspects of the review process.

Asking the right question

It is essential that the review question is right as it dictates the remit of the review. Conventionally, systematic review guidance has recommended that the review question be kept tight. It should define the intervention of interest, the population receiving the intervention, the outcome(s) of interest, and the study designs deemed worthy of inclusion.11 However, the use of systematic reviews in evaluating public health and social interventions has highlighted problems with this approach as policy or practice questions may well be broader and the answer may cover more than one intervention.21 For example, in the welfare to work review (box 1) the policy (or intervention) under review was ‘welfare to work’; however, this covers a multitude of different types of intervention (such as antidiscrimination legislation, vocational rehabilitation, return to work credits, etc), which may operate on the outcome in different ways.14 This was also the case in the reviews that examined the health effects of changing the organisation of the work environment in which task restructuring interventions ranged from the introduction of team working to increased task variety.15 Broader review questions can therefore enable the mapping of different interventions related to the overall review question.20 The use of a broad question may therefore be essential and advantageous in certain circumstances. However, the use of a broader review question is not without problems as it increases the breadth and size of the review, and potentially makes it harder to complete. It is therefore probably advantageous to break a broad review question down into smaller ones. A pilot scoping study that maps out the potential interventions of interest (and also pilots the search strategy) will also be a useful way of controlling and containing the review. The development of a review protocol is essential in this regard as it sets out the parameters, is an ongoing source of reference, and is therefore an essential element in managing the review process.11 12

The value of team work

One of the most daunting aspects of conducting a review is perhaps the isolation that it can engender. Attempting to conduct a review as a sole researcher or with very little preparation, help or support is not sensible. The systematic review methodology literature is full of comments about the importance of using a second reviewer to select and critically appraise studies independently, as well as check data extraction.11 29 Experience suggests that this is not just beneficial in terms of methodological rigour and preventing bias, but also in terms of sharing the workload and ensuring that there is support throughout the review process. Ideally, a systematic review should be conducted by a team that includes an experienced reviewer, a subject specialist and a librarian or search expert. If this is not possible then the Cochrane model of a review advisory group should be considered. Systematic reviews that use the skills of specialists are much easier to conduct and complete. The use of a search specialist has particular advantages in social science research and in terms of ensuring that a specific and sensitive search strategy is developed.30 This can save time later in the process, particularly in terms of sifting titles and abstracts. A subject specialist will also be able to assist in study location (eg, by pointing out grey literature sources or specialist databases); however, the real benefit will come in terms of the interpretation and synthesis of the results. For example, the systematic review of partnership working (box 1) used the combined skills of three researchers: a librarian, an experienced systematic reviewer and a subject expert.20
Searches: breadth versus depth

There is an implicit debate in the systematic review literature about the breadth of searches required, with some sources (purists) suggesting that only reviews that include a comprehensive search of all available evidence (via extensive searches of multiple databases, hand searches, etc.), can be called a systematic review, while others (the pragmatists) are prepared to acknowledge that it is possible to conduct a systematic review of some of the evidence base (eg, it is possible to conduct a systematic review of only one database—albeit accompanied by supplemental searches, see below) as long as this is done systematically and with transparency over the methods used and acknowledgement of the systematic review’s limitations. ‘Pragmatist’ systematic reviews therefore focus on a handful of ‘first-line’ health and social science databases, or supplement this with the use of a subject specialist one, while ‘purist’ systematic reviews have tended to search every available and potentially relevant electronic database.

Each approach has associated costs and benefits. So, for example, systematic reviews that look only at a few prominent databases save time (both at the searching stage and the sifting of titles and abstracts stage) but risk missing potentially relevant studies. Conversely, searching every database known has a high time cost, although additional studies may be located (such studies may, however, be lower down the evidence hierarchy—see below). There is evidence to suggest that combining first-line searches with a more specialist database might be the most fruitful approach as a paper reflecting on the search strategies used in a systematic review, of the effectiveness of interventions in promoting a population shift from using cars towards walking and cycling found that if only the best quality studies were included some types of intervention would not have been identified or would not have had any evidence attached to them. On the other hand, inclusive searching may mean that for well evaluated intervention types, too many lower level studies are located and they add very little additional information to the evidence base. Often when such reviews are written up it is the better quality studies that are focused on in the findings, and the results from lower level studies are marginalised. For example, the review of the compressed working week included 40 studies, yet the write-up focused only on the five prospective controlled studies. The broad study inclusion criteria on this occasion therefore increased the breadth of the review and increased the associated time and other costs (data extraction and critical appraisal are very time consuming even for experienced researchers) but with very little by way of additional information.

There is not an easy solution, but experience suggests that a trade-off is therefore required. Systematic reviews are interested in locating and synthesising the ‘best available evidence’ (not all available evidence); this means that the hierarchy of evidence does need to be applied—albeit in a pragmatic way. The best course therefore might be initially to search for all study designs so that interventions are mapped against the evidence base, but then when the parameters are known, it is only the better quality studies for each type that are subjected to the lengthy data extraction and critical appraisal process. This will therefore produce an account of the ‘best available evidence’ for each intervention type in the systematic review. For example, in

The choice of search strategy is therefore a careful balancing act and one that requires awareness of the subject and an informed judgement of the likely results alongside an assessment of resource allocation. A search strategy would therefore often benefit from an initial pilot search exercise that would produce an initial impression of where studies are located. It is not, however, a necessity—or indeed always that productive—for a systematic review to search everywhere.

The best available evidence

The transference of the systematic review methodology from evidence-based medicine was accompanied by a discussion about its limitations in respect of the very different and diverse public health and social science evidence base. Much of this focused on the applicability of the hierarchy of evidence with its emphasis on the randomised controlled trial and experimental designs. Unadulterated attempts to transfer this part of the systematic review method often led to reviews with no or uncertain conclusions. Public health systematic reviews of interventions are much more likely to be evaluated using observational and other study designs. The hierarchy of evidence, it was suggested, is therefore not useful, different approaches (such as typologies) should be utilised and public health research should look for the best available evidence wherever this may be found. This perspective is not without merit, particularly in terms of the importance of observational studies and qualitative research in evaluating public health and social science interventions. However, in practice, it has led to the implicit abandonment of any limitation in terms of study design, resulting in very broad inclusion criteria.

Broad study design inclusion criteria have merit in terms of mapping interventions. For example, an examination of the study designs of included studies in a systematic review of the effectiveness of interventions in promoting a population shift from using cars towards walking and cycling found that if only the better quality study designs were included some types of intervention would not have been identified or would not have had any evidence attached to them. On the other hand, inclusive searching may mean that for well evaluated intervention types, too many lower level studies are located and they add very little additional information to the evidence base. Often when such reviews are written up it is the better quality studies that are focused on in the findings, and the results from lower level studies are marginalised. For example, the review of the compressed working week included 40 studies, yet the write-up focused only on the five prospective controlled studies. The broad study inclusion criteria on this occasion therefore increased the breadth of the review and increased the associated time and other costs (data extraction and critical appraisal are very time consuming even for experienced researchers) but with very little by way of additional information.
the case of the compressed working week intervention this would mean controlled prospective cohorts only,18 whereas for the health walk intervention in the systematic review of transport shifts,28 it would mean potentially going as low in the hierarchy as uncontrolled retrospective cohort level evidence. The best available evidence may well vary by intervention and this is part of the mapping. However, it is important not to spend unnecessary time on the data extraction and critical appraisal of studies that do not constitute the best available evidence for any given intervention.

Tools for reviewers
Systematic reviews are very labour intensive and so it is important to avoid unnecessary replication of effort. The popularity of systematic reviews not only in evidence-based medicine but also increasingly in public health and social science means that there is the opportunity to use pre-existing and validated data extraction and critical appraisal tools. Designing, developing and piloting a new data extraction and/or critical appraisal tool is time consuming and it may also be challenged by peer reviewers when the review is completed and submitted for publication.22 There are now, however, numerous data extraction and critical appraisal tools that have already been used successfully by reviewers. These should be examined and when possible adapted for use in any new systematic review. For example, the Cochrane Reviewers Handbook or the NHS Centre for Reviews and Dissemination handbook both contain examples of data extraction forms.11 12 In terms of critical appraisal, there are various tools that can be used, the validity of which has been extensively examined. For example, the Newcastle–Ottawa scale for the assessment of observational studies has been extremely examined. For example, the Newcastle–Ottawa scale for the assessment of observational studies has attracted a lot of support and there are also now fairly well used ways of appraising qualitative evidence.39 40 Another approach would be to examine the tools used by existing systematic reviews in a similar subject area and adapt them. For example, the same data extraction and critical appraisal tools were used in all four of the systematic reviews on the work environment and they were also adapted for use in later reviews.13–20 The systematic review process is thereby considerably simplified and streamlined with valuable time saved for use in the synthesis and analysis stages.

CONCLUSION
This guide has demonstrated ways in which the ‘real’ world practice of conducting systematic reviews can be made shorter and smarter, enhancing the speed of production of systematic reviews and reducing labour intensity while still maintaining high methodological standards. It is not recommending an ‘anything goes’ approach—what is pragmatic and feasible needs to be weighed up alongside considerations of what is robust and appropriate to help ensure that systematic review findings are useful and not misleading. There are clear advantages in conducting the high quality pragmatic reviews that this guide has described: (1) time and labour resources are saved; (2) it enables reviewers to inform or respond to developments in policy and practice in a more timely manner; and (3) it encourages researchers to conduct systematic reviews before embarking on primary research, thereby reducing replication and helping to ensure that any subsequent primary research undertaken is well informed. Hopefully this guide has quelled some of the concerns of novices, elaborated on debates, and opened up the systematic review to new audiences so that it will be used more often, by more researchers, policy makers and practitioners and that it will continue to be a central part of the public health researchers’ tool box.41

What this study adds
▶ This study draws on extensive practical experience of the conduct of various ‘real’ systematic reviews across different aspects of public health policy.
▶ It highlights pragmatic ways in which systematic reviews of public health policy interventions can be undertaken without compromising their methodological integrity.
▶ It thereby dispels concerns about the systematic review and opens the method up to newcomers.

REFERENCES


