

The role of systematic reviews of qualitative evidence in evaluating interventions: A case study

Corresponding author: Theo Lorenc

Dept of Social and Environmental Health, London School of Hygiene & Tropical Medicine,
15-17 Tavistock Place, London WC1H 9SH, UK

e-mail: theo.lorenc@lshtm.ac.uk

tel: +44 (0)20 79272778

Mark Pearson

Peninsula Technology Assessment Group (PenTAG), Peninsula College of Medicine &
Dentistry, University of Exeter, UK

Farah Jamal

Institute for Health & Human Development, University of East London, UK

Chris Cooper

Peninsula Technology Assessment Group (PenTAG), Peninsula College of Medicine &
Dentistry, University of Exeter, UK

Ruth Garside

Peninsula Technology Assessment Group (PenTAG), Peninsula College of Medicine &
Dentistry, University of Exeter, UK

Abstract

Systematic reviews of qualitative evidence have been widely used to provide information on the context and implementation of interventions, and their potential barriers and facilitators. However, such reviews face a number of methodological challenges, and there are ongoing debates as to how qualitative data can best be used to inform our understanding of interventions. In this paper we use a case study of two systematic reviews of qualitative evidence on the prevention of skin cancer to explore these issues. We find that qualitative evidence not directly related to interventions is likely to be of value for such reviews; that it is often not possible to construct fully comprehensive search strategies; and that there are diminishing returns to the synthesis, in terms of added value or insight, from the inclusion of large numbers of primary studies. We conclude that there are a number of ways in which systematic reviews of qualitative evidence can be utilised in conjunction with evidence on intervention effectiveness, without compromising the rigour of the review process. In particular, the use of theory to inform frameworks for synthesis is a promising way to integrate a broader range of qualitative data.

Acknowledgements

The original work on which this paper was based was funded by the Centre for Public Health Excellence, National Institute for Health and Clinical Excellence (NICE), UK. The analyses described in the paper did not form part of the originally commissioned project. The views expressed are those of the authors and do not necessarily reflect those of NICE, or any guidance issued by NICE.

Keywords

Qualitative research; systematic review; interventions; research synthesis; comprehensiveness

The role of systematic reviews of qualitative evidence in evaluating interventions: A case study

Introduction

In recent years, systematic reviews (SRs) of qualitative evidence have become a well-established field of evidence synthesis. Early work on SRs of qualitative evidence,¹ beginning in the late 1990s, faced a number of challenges relating to search strategy development, quality assessment, and synthesis (Dixon-Woods and Fitzpatrick, 2001). While not all of these challenges have been entirely overcome, considerable progress has been made in all these areas. There has also been debate about the appropriateness of ‘conventional’ SR methodology as developed for reviews of intervention effectiveness - defined in terms of *a priori* search strategies and inclusion criteria, and a non-iterative flow of data through the review - to qualitative evidence (Dixon-Woods et al., 2006). These broader debates aside, the contribution of qualitative research to evidence synthesis is now generally recognised. SRs have been conducted across a wide range of topic areas, and have been shown to make a valuable contribution to policy and practice.

Our focus in this paper is on the use of SRs of qualitative evidence in conjunction with evidence on the effectiveness and cost-effectiveness of interventions. Such a role for SRs of qualitative evidence has been recommended on several grounds, the main one being that by understanding the barriers and facilitators of intervention effectiveness, we can ‘open up the black box’ of outcome evaluation, build more adequate theories of intervention effectiveness, and provide more useful information to assist practitioners and policy-makers in implementing interventions (Anderson et al., 2011; Noyes et al., 2011; Oliver et al., 2005; Popay et al., 1998). Other potential benefits of the use of qualitative SRs in evaluating interventions may include: explaining heterogeneity in the findings of quantitative evaluation research (Thomas et al., 2004); providing indicative *prima facie* evidence of the likelihood of intervention effectiveness in areas where quantitative outcome evaluations have not been conducted, or are methodologically inadequate (Whitehead et al., 2004); enabling researchers to engage a wider range of stakeholders by giving a richer account of the available evidence (Popay et al., 1998); and facilitating a greater understanding of the views of those affected by interventions (Harden et al., 2004).

Of course, SRs of qualitative evidence may be valuable for many reasons other than their contribution to our understanding of intervention effectiveness, and many such reviews have been conducted as stand-alone secondary research projects. However, to date, most SRs of qualitative evidence have been closely linked to questions of intervention effectiveness, either explicitly in being conducted in parallel with a systematic review of effectiveness data, or implicitly in their use of frameworks such as the barriers-facilitators schema, which involve

¹ In this paper, we use ‘qualitative evidence’, ‘qualitative data’ and ‘(primary) qualitative studies’ as synonyms, since our concern is with synthesis rather than primary research (in which the distinction between data and research studies is more relevant).

re-coding qualitative findings in terms of their potential role as a mediator of intervention success. For example, reviews of qualitative research registered with the Cochrane library are required to be linked directly to reviews of intervention effectiveness, and many policy bodies who commission and fund SRs of qualitative evidence (such as the Centre for Public Health Excellence at the UK's National Institute for Health and Clinical Excellence (NICE), who funded the work on which this paper is based) have focused primarily on using them to inform and supplement reviews of intervention effectiveness.

There are good practical and methodological reasons why this should be the case. However, it means that most SRs of qualitative evidence are conducted within a policy and practice context in which understanding intervention effectiveness is the primary goal, and understanding the public's views and attitudes mainly a means to this end. The aim of this paper is to explore the implications of this context for the conceptualisation and conduct of SRs of qualitative evidence, by comparing two reviews with a similar topic area but different policy foci.

We will not here directly address either of the two areas which include the majority of the methodological literature on SRs of qualitative evidence, namely: the debate, mentioned above, concerning the applicability of systematic review methods to qualitative evidence; and the development of methods for 'mixed-methods synthesis' to integrate the findings of SRs of qualitative evidence with those of SRs of the effectiveness of interventions. Nonetheless, our conclusions are of some relevance to both of these questions, and we hope that they may usefully inform these debates.

We draw on a case study of two SRs of qualitative evidence, led by the authors of this paper, commissioned by the National Institute for Health and Clinical Excellence (NICE) to inform the development of their guidance on the primary prevention of skin cancer (Garside et al., 2009b; Lorenc et al., 2010). The interest of this case study is that the two reviews of effectiveness focused on different sets of strategies to address the same outcomes (sun exposure and sun protection behaviours), with the corresponding qualitative SRs including relevant qualitative evidence. The similarity between the two phases of the project thus allows us to isolate the specific contribution made by each review, in the context of evaluating interventions with the aim of producing guidance for policy and practice. It enables a detailed exploration of how evidence on the public's attitudes and perceptions may be utilised within the context of the evaluation of interventions in public health.

Comparison of the phase 1 and phase 2 reviews

The guidance development process of which these reviews formed a part was split into two phases, defined by the type of interventions included. Phase 1 included information and educational interventions, while phase 2 included resource provision (such as providing free sunscreen or protective clothing), environmental change (such as the construction of shade structures) and multi-component interventions including resource provision or environmental change together with an educational component. In each phase, an SR of effectiveness and cost-effectiveness evidence and an SR of qualitative evidence were conducted. The SRs of

effectiveness for each phase looked at the same outcomes (incidence of skin cancer, sun protection behaviours, and knowledge or attitudes regarding skin cancer or sun protection) but were distinguished by the type of intervention included. The SRs of qualitative evidence for each phase were intended to locate and synthesise relevant evidence on the public's attitudes and beliefs in order to understand the potential barriers and facilitators to the success of interventions. Both reviews used the Health Belief Model (HBM) as a framework for analysis; they might therefore be considered 'framework syntheses' (cf. Carroll et al., 2011; Lorenc et al., 2008), although the framework was drawn from the included studies rather than fixed *a priori*.

Table 1 provides an overview of the two qualitative reviews, showing their inclusion criteria, outline search strategies, number of included studies and a brief indicative summary of the main findings.

[INSERT TABLE 1 ABOUT HERE]

Discussion

Overall, it can be seen from Table 1 that while the sets of papers identified and included in each review were far from identical, the findings of each are substantially similar. (A fuller description of the findings of the two reviews, and a systematic approach to translating between them, would show even greater convergence between the two.) Of course, there are some differences of emphasis, and some aspects of the analysis are unique to one review – for example phase 1's findings on public preferences regarding media messages, and phase 2's on the needs of outdoor workers – but the main themes of the synthesis, and the conclusions of the reviews, are congruent between the two. This is true not only for the more specific barriers and facilitators (e.g. the perceived severity of skin cancer) but also for the broader social and cultural meanings which were identified as themes from the qualitative evidence (e.g. aesthetic preferences for a tanned appearance, the association between tanning and health, or the interaction of sun protection behaviours with intra-family dynamics).

Our discussion of these findings focuses on three methodological areas:

- the inclusion criteria to be employed for SRs of qualitative evidence in the context of evaluating interventions;
- the development of search strategies, and the issue of congruence between search strategies and inclusion criteria; and
- methods for the synthesis of qualitative evidence.

Finally, we consider the broader implications of our findings, and offer a few suggestions for further work in the methodology of SRs of qualitative evidence.

Inclusion criteria

For the phase 1 review, the inclusion criteria were initially interpreted so as to restrict inclusion to only those studies which presented qualitative evidence relating to the evaluation of a specific skin cancer prevention intervention. However, this criterion was dropped early in the review, since it proved to be overly restrictive: only four of the 16 reports which were finally included presented such data. Moreover, it was clear that useful and relevant data were to be found in studies which investigated attitudes to sun behaviours or skin cancer in general, without reference to interventions. Similar criteria were adopted for the phase 2 review at the outset, on the basis of the experience of the phase 1 review team. Again, few studies of interventions were included in the phase 2 review (three of 23 reports). Hence, for both reviews, the inclusion criteria were finally interpreted so as to include any study reporting qualitative evidence relating to sun protection beliefs or behaviours, regardless of its link to a specific intervention, since it was assumed that these would potentially indicate areas of resistance to or support for adopting safe sun behaviours.

Thus, it was found to be necessary in both reviews to include evidence not relating directly to interventions. In this respect, these reviews appear to be generally representative of the field of public health, where relatively little substantive qualitative evidence on specific interventions is available. Indeed, this is probably the case in many areas of social and health research. If so, limiting inclusion to qualitative studies of interventions alone will not be a practicable course of action, due to the lack of data (cf. Garside et al., 2009a).

Moreover, our findings indicate that data relating to specific interventions are generally not, in fact, of greater value than data which relate to broader attitudes. Although some studies of specific interventions were located in both reviews, their contribution to the synthesis was relatively limited, and studies not relating to interventions were more numerous and generally more useful. Thus, as previous methodological studies have found, qualitative data on the implementation of specific interventions may be of limited value (Roen et al., 2006). This empirical finding supports the *a priori* point that, to the extent that SRs of qualitative evidence aim to access broader perspectives and contexts, including those which do not relate directly to interventions, there are good reasons not to limit inclusion to studies of interventions. In our view, this raises broader questions about the role of qualitative evidence, which are explored further below.

A corollary is that, even where SRs of qualitative evidence are conducted in parallel with reviews of intervention effectiveness, they will often need to adopt different conceptual schemata for their inclusion criteria and search strategies. Existing methodological guidance indicates that the schema used for the effectiveness review, for example PICO (population, intervention, comparator, outcomes), should be adapted for the SR of qualitative evidence so as to maintain a similarity of structure between the two (NICE, 2009, p.48; Ring et al., 2011, p.9). However, the considerations above suggest that this may be inappropriate in many cases, and that a degree of structural divergence between SRs of qualitative evidence and SRs of effectiveness may be inevitable.

Searching

As already noted, despite the very similar criteria used for the two reviews, a substantial number of study reports were included in one but not the other (a total of 21 out of 30 across the two reviews). The large number of non-overlapping studies raises potentially troubling questions about the comprehensiveness and coherence of the reviews, since many of the studies not located by each review would have met the criteria if they had been returned by the searches. The reviews therefore cannot be said to have been comprehensive, in the sense of locating all available studies meeting the inclusion criteria.²

This non-comprehensiveness is primarily due to the search terms relating to interventions which were employed in the search strategies. Since, as discussed above, inclusion was not restricted to studies relating directly to the evaluation of specific interventions, the search strategies and inclusion criteria were not precisely congruent. In effect, these intervention search clusters acted as a filter which - although maintaining the transparency of the search process - excluded a substantial amount of potentially relevant evidence. Thus, tying the search terms of a SR of qualitative evidence too closely to interventions may compromise the consistency of the review. More generally, this finding suggests that methodological decisions about the conduct of such reviews cannot be straightforwardly derived from research questions focused on intervention effectiveness.

This issue is generally recognised. As the *Cochrane Handbook* (Noyes et al., 2011, section 2.3) puts it:

Qualitative evidence retrieved using a topic-based search strategy designed to identify trials cannot be viewed as being either comprehensive or representative. Such a search strategy is not designed for the purpose of identifying qualitative studies and indeed achieves a measure of specificity by purposefully excluding many qualitative research types.

However, the implications of such exclusion have been less widely discussed. Our findings might lead us to question whether intervention clusters should be used in qualitative SRs at all, if (as we have argued is generally the case) such reviews cannot be restricted to qualitative studies of interventions alone. In this case, the goal of comprehensiveness would demand that these restrictive intervention terms be dropped. However, if this were done with no other change to the strategies adopted in these reviews, the strategies would become highly over-inclusive and the volumes of records impracticably large. (This also prevents us from precisely quantifying the impact of the intervention terms, since to do so, we would need to re-screen all the results of these vastly more inclusive searches.)

It is not clear that there is any satisfying solution to this issue. The difficulty of creating clusters of methodological terms to locate qualitative research with a high degree of specificity is well-known (Evans, 2002; Grant, 2004), and, in the case of our reviews, there

² We might call this 'numerical' or 'categorical' comprehensiveness; the broader notion of theoretical comprehensiveness, in the sense of addressing all the contexts important for the intervention of interest, is not directly at issue here, although cf. our remarks about the concept of saturation below.

was no clear rationale for restricting inclusion, for example by population or date, in order to reduce volume. While our reviews may not be representative of all SRs of qualitative evidence in this respect - in particular, reviews where the focus of the intervention is narrowly defined may face less of a challenge in terms of volume - similar issues are likely to arise with respect to many policy- and practice-relevant questions.

Our findings thus support the view that comprehensiveness is in many cases not an attainable goal for SRs of qualitative evidence, particularly in the context of evaluating effectiveness. This is probably not as serious a limitation as it seems, since, as we go on to argue in the following section, comprehensiveness may not be a desirable goal for such reviews either. Nonetheless, it does imply that that, in many cases, the search strategies for such reviews will not practically be able to cover the whole scope of the review as defined by the inclusion criteria, but will require some restriction. In the case of the reviews discussed here, this restriction took the form of a cluster of terms for interventions. However, other ways of approaching the problem – perhaps by focusing on particular theoretical constructs or methodological approaches – would be equally legitimate in principle (see further under ‘Broader implications’ below).

Synthesis

As already noted, the high-level findings of the two reviews were largely identical, despite the substantially different data sets on which they are based. While there are some divergences of interpretation, the main messages are largely congruent and in some cases very similar indeed. While several primary study reports (N=9) were included in both reviews, the thematic congruence does not appear to be mainly driven by these overlapping studies, but emerges equally from the studies unique to each review (with minor exceptions,³ no thematic area in either review relied entirely on the overlapping studies).

The similarity of conclusions between the two reviews speaks for the reliability of the synthesis process, in that a substantially similar end-point was reached by different research teams, working largely independently (although the phase 2 review team were aware of the phase 1 review). However, it also suggests that, for these reviews, there were rapidly diminishing returns in terms of extra insight or validity from the inclusion of larger numbers of studies, since the broad outline of the conclusions could have been reached with a smaller body of research than that in either review. For example, the main findings regarding the low perceived salience of sun protection, or the preference for a tanned appearance, could have been gained from a relatively limited overview of the literature. This is partially due to the use of the HBM as a synthesis framework common to both reviews. Nonetheless, it seems likely that these diminishing returns would have been observed even with formally distinct frameworks, or the use of grounded-theory methods. Thus, we would suggest that this finding is relevant to syntheses of qualitative evidence in general.

³ Some of the findings about schools and recreation settings rely on single studies which were included in both reviews.

The finding that increasing numbers of studies produces little extra value might seem to support the proposal that the concept of ‘saturation’, familiar from primary qualitative research and particularly the grounded theory literature (Glaser and Strauss, 1967), could be used to guide study selection for reviews (Mays et al., 2005). As Booth (2001) incisively puts it:

Why should systematic reviewers of qualitative research pursue a "gold standard" comprehensive literature search when concepts such as "data saturation" have an established pedigree? ... Interestingly quantitative reviewers are currently seeking methods to define a 'law of diminishing returns' beyond which further literature searching has little justification in order to manage the inordinate expense of the searching process. For qualitative reviews the answer to this problem already exists in the principles of data saturation used in primary studies.

Our findings indicate that saturation was reached at a relatively early stage in the reviews, and hence that using the principle of saturation to guide the review could have led to considerable gains in efficiency. However, although this idea is theoretically appealing, it is as yet unclear how it might be implemented (Thomas and Harden, 2008).

One method would be to base the inclusion process on the principle of saturation such that, rather than including all studies meeting the criteria, a selection could be made (based on the perceived value of the primary studies, or on *a priori* theoretical grounds), and inclusion extended to new studies up to the point where they no longer add to the content of the synthesis. Such an approach has similarities to realist review methodology, where papers are selected based on rigour and relevance to the research question (Pawson et al., 2005). However, it remains an open question whether any saturation-based screening method can be implemented in such a way as to maintain the transparency and reproducibility of the SR process.

An alternative would be to operationalise the principle of saturation at the level of the synthesis, rather than the level of study inclusion. For example, in meta-ethnography, the conceptually richer papers have greater weight in the synthesis, with other papers merely illustrating these concepts with descriptive themes (Britten et al., 2002; Garside et al., 2008). However, to the extent that the process underlying such synthesis methods remains that of a traditional SR, the gains in the efficiency or scope of the review as a whole are likely to be small (although the gains in terms of insight may be substantial).

Moreover, it is unclear that the concept of ‘saturation’ can be directly transferred from primary to secondary research. Many of the techniques available to primary researchers to test for saturation are not applicable to reviews. For example, a primary researcher may purposively sample cases which are likely to disconfirm their findings, in order to confirm that saturation has been reached. For secondary researchers, this will be possible to a much more limited extent, since they are dependent on the available literature: if no studies have been conducted which illuminate a particular dimension of saturation (e.g. which include a particular population group), such purposive testing of saturation will not be possible. Thus,

our findings, rather than showing that saturation was reached early on in the reviews, may indicate the difficulty of robustly testing whether saturation has in fact been achieved, and hence the inadequacy of saturation as a principle of synthesis.

Because of this, in practice, identifying the point of saturation in a review of qualitative evidence will often be determined primarily by the *a priori* methodology and objectives of the review, rather than by the data themselves, as should ideally be the case in a primary study. Indeed, the level at which the concept of saturation should be applied will largely depend on the aims and context of the review. A review which, like the ones examined here, seeks primarily to draw together and summarise what is known about the public's views and attitudes, and strongly emphasises the identification of themes which are common to several primary studies, will aim mainly for *thematic* saturation, and, if our findings are any guide, is likely to reach it at an early stage. However, the literature on the synthesis of qualitative evidence (and that on primary qualitative research methods) has not always clearly distinguished thematic saturation from *theoretical* or *conceptual* saturation, which forms the main focus of grounded theory in its original form (Glaser and Holton, 2004). This latter concept might be of greater importance in reviews which take a more interpretive or theory-led approach. (For example, a theory-led approach to the data in our reviews might have drawn on the Foucauldian concept of the 'disciplinary gaze' employed by one of the included primary studies (Carter, 1997) to develop a more critical account of the relation between sun protection behaviours and attitudes and health promotion agendas.) Conceptual saturation in this sense may take considerably longer to attain than thematic saturation; it is also likely to involve bodies of theory and data outside the review itself, and to depend more fundamentally on the broader goals of the review, and the research programme or policy-making process of which it forms a part.

We return to the question of different levels of synthesis in the conclusion. The immediate point is that, while our findings provide strong *prima facie* evidence for the importance of saturation, this concept needs to be addressed more critically as a potential basis for qualitative syntheses, particularly where it is taken to mean thematic saturation in isolation from conceptual saturation. A strongly saturation-based approach runs two risks: first, limiting saturation to "simple redundancy without conceptual analysis" (Glaser and Holton, 2004) and hence prioritising the aggregation of thematic content over higher-level theory-building; and second, introducing bias into the conclusions of the review by presenting methodological choices (e.g. a focus on barriers and facilitators of interventions) as substantive conclusions. An approach which admits the limitations of the synthesis process, and the possibility of alternative interpretations or theoretical perspectives, may provide a more adequate account of the available evidence.

Independently of this broader point, these findings provide further reason to question the value of comprehensiveness in SRs of qualitative research. In the previous section, we noted that comprehensiveness may not be an attainable goal for such reviews; the findings described in this section indicate that it may not be a desirable goal either, in that both our reviews adequately met the need for a robust summary of relevant qualitative data without being comprehensive in a formal sense. More generally, we might be led to question the

assumption implicit in some methodological work that more is necessarily better. It appears unlikely that a more comprehensive review – to the extent that this is possible – would substantially add to the findings, although it would doubtless add useful detail on certain points. This said, we cannot conclusively demonstrate that this is the case without extensive further searching. In particular, it is impossible to know whether a fully comprehensive review would have found studies which could inform substantively different synthetic constructs. These reservations aside, our findings provide a clear indication that comprehensiveness in qualitative reviews is likely to have limited impact at the level of substantive findings.

Finally, our findings also suggest that researchers undertaking primary qualitative studies would benefit from SRs of the literature in the area in which they plan to research, since the studies included in our reviews overlapped substantially in terms of methods, research questions and populations, and many did not reference similar work undertaken previously. As in other fields, SRs of qualitative evidence can help to guide the planning of research to maximise its value and impact.

Broader implications

Our findings suggest that the use of qualitative SRs in the context of evaluating interventions faces certain challenges. However, we do not think that these concerns provide grounds for rejecting qualitative SR as a methodology. On the contrary, the reviews, and the guidance which they informed, clearly bear out the value of qualitative evidence as part of evidence synthesis designed to inform policy and practice. We hope that these findings may contribute to a broader debate about how qualitative SRs might be best used in this context.

We can begin to outline the scope of such a debate by observing that the process of any synthesis of qualitative evidence involves drawing the primary studies into a dialogue across differences of population, context, and in some cases, methodology or research paradigm. Even when the data are formally homogenous, producing a coherent synthesis implies the recognition that data from one context can be seen as relevant to data from other contexts in various different ways, which will often not be specifiable in advance of the synthesis itself. The choice of which direction to follow in the synthesis may be constrained on pragmatic grounds – for example, by the use of an *a priori* framework such as the Health Belief Model – but other interpretive options always remain available.

This point remains true when we consider the relevance of qualitative studies not only to each other, but to studies of intervention effectiveness. That is, in principle, qualitative evidence may inform and support our understanding of intervention effectiveness in many different ways. While it is likely that the majority of qualitative SRs will remain subordinate to SRs of effectiveness for the foreseeable future - in the sense that questions about effectiveness will continue to be the primary influences shaping broader research agendas, and determining the

relevance of qualitative data - this leaves open the question of *how* the relevance of qualitative data is determined in particular cases.⁴

This conclusion indicates that more flexible and pluralistic approaches to the use of qualitative evidence may be viable. These might take a range of forms, depending on the context and the aims of the research. One idea here is to look at a wider set of dimensions in the quantitative evidence which could be addressed by qualitative data. The widespread use of the barriers-facilitators framework, or similar schemata for synthesis such as the Health Belief Model, tends to limit attention to the desired outcomes of interventions: that is, qualitative evidence is considered relevant only if it directly addresses the outcomes which are the focus of the review of effectiveness. However, other levels and dimensions of the intervention research might equally be taken as a focus, for example: the population or setting (e.g. schools, workplaces, the internet); the specific intervention strategies employed, and the experiences of those involved in planning or delivering them; or the organisational-level barriers and facilitators of successful intervention campaigns or policy-level strategies. In each case, the scope of relevant qualitative evidence, and the methods for locating and synthesising it, would be different. NICE's own guidance on behaviour change (<http://guidance.nice.org.uk/PH6>), which draws together evidence on a wide range of interventions, might suggest useful avenues of investigation here. More broadly, we might seek to go beyond such categorical frameworks for synthesis to more substantively theory-led approaches, which seek to develop an integrated understanding of the whole field of interest, in order to facilitate a more explicit and wider-ranging consideration of the ways in which qualitative research may be valuable in understanding intervention effectiveness.

Finally, we would like to suggest some potential directions for future methodological research. One idea here is that SRs could draw on the insights of other evidence synthesis methodologies regarding the benefits of iterative and theory-led approaches, whereby emerging constructs inform the selection of data for future stages of the review. In the SR context, this would involve substantially elongating the initial scoping stages of the review before a final search strategy and inclusion criteria were finalised. While this may have considerable resource implications, some practicable methodological options are available. For example, reviews can include a dedicated theory-building stage at the beginning of the process, which can be relatively brief and pragmatic in its methodology (Lorenc et al., 2011). More speculatively, the growing viability of text-mining techniques may provide ways to make this process more efficient, by using techniques such as automated document clustering to provide an initial overview of the available evidence base across a broad range of topic areas (Ananadiou et al., 2009; Thomas et al., 2011). However, our findings indicate that such

⁴ Although methods for reviews of effectiveness are beyond the scope of this paper, this conclusion might also suggest a need to think more critically about the concepts of 'intervention' and 'effectiveness' themselves, and about whether the focus of evaluation should be interventions or the theories which underlie them.

methodological innovations will need to be placed within a broader process of reflection on the utility and value of qualitative reviews, and their role in providing evidence on interventions.

Conclusion

The value of SRs of qualitative evidence in conjunction with reviews of effectiveness is increasingly recognised. However, there are a number of outstanding questions about the appropriate methodology for such reviews. Our analysis of two SRs of qualitative evidence on closely related topics raises three issues: first, the scope of such reviews as fixed by the inclusion criteria will usually need to be broad, including evidence not directly related to interventions; second, due to this breadth, search strategies will usually not be able to attempt comprehensiveness with respect to the full scope of the review, since this would result in an impracticable volume of returns; and, third, with the inclusion of greater numbers of studies, their added value appears to diminish quite rapidly, at least within a given synthesis framework.

These findings do not support any generalised scepticism about SRs of qualitative research. However, they do suggest that such reviews are, to some extent, methodologically *sui generis*, and cannot be governed solely by concepts imported either from SRs of quantitative evidence (e.g. comprehensiveness) or from primary qualitative research (e.g. saturation). Finally, where SRs of qualitative evidence and of effectiveness are undertaken on the same topic in parallel, our findings indicate that it may be counter-productive to attempt to maintain a strict isomorphism between them, and to minimise any potentially contestable theoretical assumptions. More flexible and theoretically informed approaches may be equally robust in practice, while offering greater explanatory power.

References

- Ananadiou S, Procter R, Rea B, Sasaki Y, Thomas J (2009) Supporting systematic reviews using text mining. *Soc Sci Comput Rev* 27:509-523.
- Anderson LM, Petticrew M, Rehfuss E, Armstrong R, Ueffing E, Baker P, Francis D, Tugwell P (2011) Using logic models to capture complexity in systematic reviews. *Research Synthesis Methods* 2:33-42.
- Booth A (2001) Cochrane or cock-eyed? How should we conduct systematic reviews of qualitative research? Paper presented at the Qualitative Evidence-based Practice Conference, Taking a Critical Stance. Coventry University, 14-16 May 2001. Available at: <http://www.leeds.ac.uk/educol/documents/00001724.htm>. Accessed 27 October, 2011.
- Britten N, Campbell R, Pope C, Donovan J, Morgan M, Pill R (2002) Using meta-ethnography to synthesise qualitative research: a worked example. *J Health Serv Res Policy* 7:209-215.
- Carroll C, Booth A, Cooper K (2011) A worked example of "best fit" framework synthesis: A systematic review of views concerning the taking of some potential chemopreventive agents. *BMC Med Res Meth* 11:29.
- Carter S (1997) Who wants to be a "peelie wally"? Glaswegian tourists' attitudes to sun tans and sun exposures, in *Tourism and Health* (Clift SGP ed), pp 139-150. Pinter, London.
- Dixon-Woods M, Bonas S, Booth A, Jones DR, Miller T, Sutton AJ, Shaw RL, Smith JA, Young B (2006) How can systematic reviews incorporate qualitative research? A critical perspective. *Qual Res* 6:27-44.
- Dixon-Woods M, Fitzpatrick R (2001) Qualitative research in systematic reviews. Has established a place for itself. *BMJ* 323:765-766.
- Evans D (2002) Database searches for qualitative research. *J Med Libr Assoc* 90:290-293.
- Garside R, Britten N, Stein K (2008) The experience of heavy menstrual bleeding: a systematic review and meta-ethnography of qualitative studies. *J Adv Nurs* 63:550-562.
- Garside R, Pearson M, Moxham T, Anderson R (2009a) Barriers to, and facilitators for, the effectiveness of multiple risk factor programmes aimed at reducing cardiovascular disease within a given population: a systematic review of qualitative research. National Institute for Health and Clinical Excellence, London. Available at: <http://www.nice.org.uk/guidance/index.jsp?action=download&o=49377>. Accessed 27 October 2011.
- Garside R, Pearson M, Moxham T, Anderson R (2009b) Providing public information to prevent skin cancer. Review 2. Barriers and facilitators to conveying information to prevent first occurrence of skin cancer: a systematic review of qualitative research. National Institute for Health and Clinical Excellence, London. Available at:

<http://www.nice.org.uk/guidance/index.jsp?action=download&o=52644>. Accessed 27 October 2011.

Glaser B, Holton J (2004) Remodeling Grounded Theory. *Forum Qual Soc Res* 5:4.

Glaser B, Strauss A (1967) *The discovery of grounded theory: Strategies for qualitative research* Chicago, Aldine.

Grant MJ (2004) How does your searching grow? A survey of search preferences and the use of optimal search strategies in the identification of qualitative research. *Health Info Libr J* 21:21-32.

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

Lorenc T, Brunton G, Oliver S, Oliver K, Oakley A (2008) Attitudes to walking and cycling among children, young people and parents: a systematic review. *J Epidemiol Community Health* 62:852-857.

Lorenc T, Clayton S, Neary D, Petticrew M, Whitehead M (2011) Causal mapping for systematic reviews of complex social interventions. Poster presented at 19th Cochrane Colloquium, 19-23 October 2011, Madrid.

Lorenc T, Jamal F, Cooper C (2010) Sun protection resources and changes to the environment to prevent skin cancer: Qualitative evidence review. National Institute for Health and Clinical Excellence, London. Available at: <http://www.nice.org.uk/guidance/index.jsp?action=download&o=52647>. Accessed 27 October 2011.

Mays N, Pope C, Popay J (2005) Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field. *J Health Serv Res Policy* 10 (Suppl. 1):6-20.

NICE (National Institute for Health and Clinical Effectiveness) (2009) *Methods for the Development of NICE Public Health Guidance*, 2nd edition. National Institute for Health and Clinical Excellence, London. Available at: www.nice.org.uk/media/2FB/53/PHMethodsManual110509.pdf . Accessed 8 December 2011.

Noyes J, Popay J, Pearson A, Hannes K, Booth A (2011) Chapter 20: Qualitative research and Cochrane reviews, in *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* (Higgins JPT, Green S eds). The Cochrane Collaboration. Available at: www.cochrane-handbook.org. Accessed 27 October 2011.

Oliver S, Harden A, Rees R, Shepherd J, Brunton G, Garcia J, Oakley A (2005) An emerging framework for including different types of evidence in systematic reviews for public policy. *Evaluation* 11:428-446.

Pawson R, Greenhalgh T, Harvey G, Walshe K (2005) Realist review--a new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy* 10 (Suppl. 1):21-34.

Popay J, Rogers A, Williams G (1998) Rationale and standards for the systematic review of qualitative literature in health services research. *Qual Health Res* 8:341-351.

Ring N, Ritchie K, Mandava L, Jepson R (2011) *A Guide to Synthesising Qualitative Research for Researchers Undertaking Health Technology Assessments and Systematic Reviews*. NHS Health Quality Improvement Scotland, Edinburgh. Available at: <http://www.nhshealthquality.org/nhsqis/8837.html>. Accessed 8 December 2011.

Roen K, Arai L, Roberts H, Popay J (2006) Extending systematic reviews to include evidence on implementation: methodological work on a review of community-based initiatives to prevent injuries. *Soc Sci Med* 63:1060-1071.

Thomas J, Harden A, Oakley A, Oliver S, Sutcliffe K, Rees R, Brunton G, Kavanagh J (2004) Integrating qualitative research with trials in systematic reviews: an example from public health. *BMJ* 328: 1010-1012.

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

Thomas J, Mcnaught J, Ananadiou S (2011) Applications of text mining within systematic reviews. *Research Synthesis Methods* 2:1-14.

Whitehead M, Petticrew M, Graham H, Macintyre SJ, Bambra C, Egan M (2004) Evidence for public health policy on inequalities: 2: assembling the evidence jigsaw. *J Epidemiol Community Health* 58:817-821.

Table 1: summary of Phase 1 and Phase 2 qualitative reviews

	Phase 1 review (Garside et al)	Phase 2 review (Lorenc et al)
Intervention review focus	Information and educational initiatives	Resource provision (such as free sunscreen) and environmental change
Inclusion criteria	Data relating to skin cancer, sun protection, sunbathing/tanning	
	Study type: Qualitative research	
	Date: 1990-present	
	Language: English	
	Country: OECD member states	
	Views relevant to information or education	Views relevant to environmental change, resource provision or multi-component interventions
Structure of database search strategy	- ((skin cancer terms) OR (sun / ultraviolet terms)) AND - (prevention / health promotion terms) AND - (education / information terms) AND AND - (qualitative methods terms)	- (skin cancer terms) AND - (sun / ultraviolet terms) AND - (prevention / health promotion terms) AND - ((environment terms) OR (resource provision terms)) AND - (qualitative methods terms)
No. included studies and overlap	16 reports of 15 studies (9 reports also in Phase 2 review)	23 reports of 22 studies (9 reports also in Phase 1 review)
Synthesis framework	Health Belief Model	
Findings: Susceptibility	Generally low perceived susceptibility	
Findings: Severity	Generally low perceived severity	
	Skin ageing perceived as equally serious as cancer	
Findings: Benefits	There was an awareness of the potential damaging effects of sun exposure but also some limited understanding of sun safety messages	Sun protection (esp. sunscreen) widely seen as preventing cancer and skin ageing
Findings: Barriers	Tans are 'healthy' and connote a physically active lifestyle	
	Tans are attractive and increase confidence and psychological well-being	
	Tans connote a good holiday	Sunscreen associated with beaches / holidays
	Peers' views influence sun behaviours	
	Practical and social barriers to sun protection	
	Structural challenges in schools	
		Being outdoors in the sun feels 'healthy'
	'Incidental' tanning not seen as calling for protection	
	Sun exposure seen as less risky than sunbeds	

	Teenagers' growing independence may compromise sun protection behaviour	Messages seen as more relevant to younger children than older children or adults
Findings: Cues to action	Parents are often a source of encouragement	Parents (esp. mothers) lead sun protection within family
	Older children listen more to peers than parents/teachers	
	Knowing people who have had cancer is a source of encouragement	
		Policies in schools may be more promising as part of 'whole school' approach
	Media campaigns are generally seen as credible, but may not be the main source of influence and are sometimes seen as simplistic	
Findings: Self-efficacy	Skin cancer is seen as preventable by individuals taking responsibility	
Findings: Other	Longer history of sun safety messages and stronger regulation in Australia than UK or Canada. UK policy does not address desirability of tanning.	
		Women more likely to use sun protection than men, but also more likely to deliberately sunbathe; this perception linked to broader gender norms
		Outdoor workers are of particular concern, and may be hard to reach