
Article:
Carmona, C. orcid.org/0000-0001-9855-3575, Baxter, S. orcid.org/0000-0002-6034-5495 and Carroll, C. orcid.org/0000-0002-6361-6182 (2021) Systematic review of the methodological literature for integrating qualitative evidence syntheses into health guideline development. Research Synthesis Methods, 12 (4). pp. 491-505. ISSN 1759-2879

https://doi.org/10.1002/jrsm.1483

Received: 10 August 2020 | Revised: 31 December 2020 | Accepted: 12 February 2021

DOI: 10.1002/jrsm.1483

RESEARCH ARTICLE

Research Synthesis Methods WILEY

Systematic review of the methodological literature for integrating qualitative evidence syntheses into health guideline development

Chris Carmona | **Susan Baxter** | **Christopher Carroll**

School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK

Correspondence
Chris Carmona, School of Health and Related Research, University of Sheffield, Sheffield, UK.
Email: ccarmona1@sheffield.ac.uk

Abstract

Guidelines produced by local, national and international bodies underpin clinical practice and healthcare services worldwide. For guidelines to be based on the best available evidence, it is critical that syntheses of both qualitative and quantitative evidence are used to inform decision-making. As methods for qualitative evidence syntheses (QES) develop, they are increasingly able to inform health guideline production. However, the process whereby this form of evidence is considered and incorporated tends to be unclear. This systematic review synthesized existing guidance concerning the use of QES in guideline development. Sources published in English that described or prescribed methods for incorporating QES into evidence-based health guidelines were eligible for inclusion. Seventeen relevant papers were identified. The literature indicates that there is a reasonable consensus about many stages of conducting a QES to inform guideline development. Areas needing further exploration include: the way that committees engage with QES; the usefulness of different QES methodologies; and understanding of how expert committees use evidence. Methods for producing QES for guideline committees tend to be similar to quantitative systematic review methods in terms of searching, quality appraisal, systematic management of data, and presentation of results. While this allows transparency and accountability, it could be argued that it is less "true" to the principles of being led by the data, which are fundamental to most qualitative research. Understanding the process of using QES to produce guidelines is critical to determining their validity and applicability, and to ensure that healthcare provision is based on the best available evidence.

KEYWORDS

health guidelines, qualitative evidence synthesis, systematic review

1 | INTRODUCTION

Guidelines produced by local, national and international bodies are used to underpin clinical practice and the delivery of healthcare services worldwide. The World Health Organization (WHO), for example, lists 239 guidelines on its website,[1] and the National Institute for Health and Care Excellence (NICE) in the United Kingdom currently lists 1623 guidance products, including 354 guidelines.[2] Understanding the process of producing these guidelines is critical to determining their validity and applicability, and to ensure that healthcare provision is based on the best available evidence.

The use of findings from qualitative evidence syntheses (QES) as evidence in the development of health guidelines is growing as the need for relevant and context-sensitive evidence increases.[3] This is commonly agreed to be because qualitative data can answer particular types of questions far better than quantitative data. Quantitative data are still key for questions of efficacy, but are less able to answer questions relating to understanding of patient preference, and other contextual outcomes such as feasibility and acceptability. These questions are best answered by qualitative studies.[4] QES might usefully answer questions that are key to guideline production, for example, how different groups of practitioners, people using services or stakeholders perceive the issue, what social and cultural beliefs, attitudes or practices might affect this issue or how different groups perceive the intervention or available options.[5]

Increased recognition of the value of QES is also driven by the move toward greater patient-centredness in health systems, for example, an emphasis on shared decision making, and the greater inclusion of patients and lay experts in guideline producing committees.[6] It is also claimed that incorporating QES into guideline development can help to represent people who may otherwise be excluded from the process,[7] and that QES can also potentially offer a valuable supplement to the experiences of patient representatives on guideline panels.[8] This does not mean that advocates of QES in guideline development are oblivious to the challenges of the approach. A number of concerns have been reported about using QES within international guideline-producing bodies. First, authors acknowledge that guideline producers are principally systematic reviewers who may have no background or expertise in qualitative methods, and therefore need training to be able to produce high quality QES.[5,9,10] Second, many qualitative researchers do not support the practice of synthesising qualitative research, and that for those that do there is no universally accepted way of doing this in health and social care[6] (though this position has changed somewhat since the publication of that paper). Standardization of methods for producing QES is contrary to many qualitative approaches that are data led and iterative. There is a call for QES not to violate the underpinning epistemological foundations of the included studies.[11] Third, qualitative research itself has been criticized by positivist authors as being context-dependent and specific, for including an insufficient number of informants, for being interpretative and, because it usually relies on small, purposive samples, for having a low degree of generalization.[12,13] Conversely, this is regarded by qualitative researchers as one of the great strengths of qualitative research. It has also been argued that there are issues with the linking, mixing or merging of qualitative and quantitative evidence, and there is no ready-made toolkit for doing this.[8] The purpose of the current study however is to explore the methods of qualitative evidence synthesis in health guidelines, not to argue for or against their use.

A review of the use of qualitative data by NICE up to 2009, in addition noted a lack of consistency in terminology and method and even lack of agreement about what constituted a qualitative study across their different guideline producing centers.[10] Other authors have highlighted a lack of clarity about the processes involved in how committees make decisions on the basis of qualitative evidence, and furthermore, how the strength of evidence relates to the strength of recommendation when QES is included.[4,7] For example, WHO guidelines have been criticized for making "strong recommendations" despite there being only low or very low confidence in the underpinning evidence.[4] However, WHO argue that their guideline panels are expected to take into account broader evidence about acceptability, feasibility, and equity, in addition to evidence about effectiveness.[4]

Methods for the synthesis of quantitative evidence are well established, and robust methods for meta-analysis and the pooling of quantitative data provide clearly interpretable information for decision making bodies. Interpretation of the available evidence is also supported by an established framework for determining its quality through use of the GRADE tool.[14]

Alongside the ongoing concerns over their use and the readiness of guideline producing bodies to integrate QES evidence into their processes, it is crucial to examine the methods that are being adopted or proposed both by experts in the field, and by guideline producing bodies themselves. Recent growth and development in methods and standards for QES, and the development of tools to ensure standardization both in QES (e.g., CERQual[15] and the work of the Cochrane qualitative and implementation methods group[16]) and in guideline development (e.g., the DECIDE collaboration evidence to decision frameworks[17,18]) has put qualitative evidence firmly on the agenda for evidence-based medicine. However, the most appropriate methods for using it during guideline

development remain unclear. This study aimed to systematically review the methodological literature that addresses this topic, produce a synthesis of the state of the field, and explore where consensus and disagreement may exist.

2 | METHODS

The review question was: what methods and processes are proposed in the methodological literature for incorporating QES into evidence-based health guideline development? As a methodological review the protocol was not eligible for PROSPERO registration, therefore it is provided as File S1.

2.1 | Eligibility criteria

Papers, online sources or published manuals that described or prescribed methods for incorporating QES into evidence-based health guidelines were eligible for inclusion if they gave sufficient detail to allow extraction of different stages of the process and the methods used in those stages. Since the papers included were descriptive and not empirical studies, no study design restrictions were placed. Similarly, no country restrictions were put in place, although included papers needed to be published in English.

2.2 | Information sources and search strategy

Health related databases were searched for papers published in English since 2000. The date was selected because neither qualitative evidence synthesis methodology, nor methods for guideline development were well established before that time. Database searches were conducted in Medline (including Medline in process), EMBASE, CINAHL and PsycINFO from 2000 up until August 15, 2019. Supplementary searches were conducted in Google Scholar and results from the first six pages were added to the search results. Reference lists of papers were checked for further potential includes. A full search history can be found in File S2.

2.3 | Data collection process

Titles and abstracts for all papers were screened, and those that appeared to meet the inclusion criteria, or those where it was uncertain whether the criteria were met or not were examined as full text. Articles marked for inclusion at this stage also had their reference lists checked to identify further papers. To maximize transparency of selection at the full text stage, a checklist was used to ensure that papers met the criteria for inclusion as described above (see File S3, for the completed checklist).

2.4 | Data extraction

An a priori data extraction framework was set up in NVivo 12[19] to map the different stages of the reviewing process. The stages were chosen to represent the potential range of discrete tasks that are involved in a quantitative systematic review or QES, that is, protocol/scope/review question; searching; study selection; data synthesis, critical appraisal; quality appraisal; making recommendations, use of logic models/frameworks; integration with quantitative data/reviews; and reporting. Three additional "umbrella" categories were also used to capture broader themes about the use of QES in guidelines - benefits of using QES in guidelines; challenges of using QES in guidelines; and QES methodologies that have been used in guideline development. Additional emergent themes were coded as they occurred during the data extraction process.

2.5 | Risk of bias/quality appraisal

Formal risk of bias or quality assessment of included papers was not appropriate as they were methodological rather than empirical studies however, the design and any key weaknesses of included papers were noted during data extraction.

2.6 | Method of synthesis of results

Data extracted was synthesized narratively, within the coding categories described above. Particular attention was paid to possible overlaps between different stages and the overarching "umbrella" categories.

3 | RESULTS

3.1 | Study selection

Searching of databases yielded a total of 5822 references. These were uploaded into EPPI reviewer 5 software[20] and de-duplicated. A total of 756 duplicate records were identified and removed. In total, 5066 records were screened at title and abstract level. 5019 records were excluded with 47 papers marked for full text examination. For a list of papers excluded at full text, along with reasons for their exclusion, see File S6. Eleven articles from data base searching and six additional references from reference list checking met the inclusion criteria. A total of 17 papers

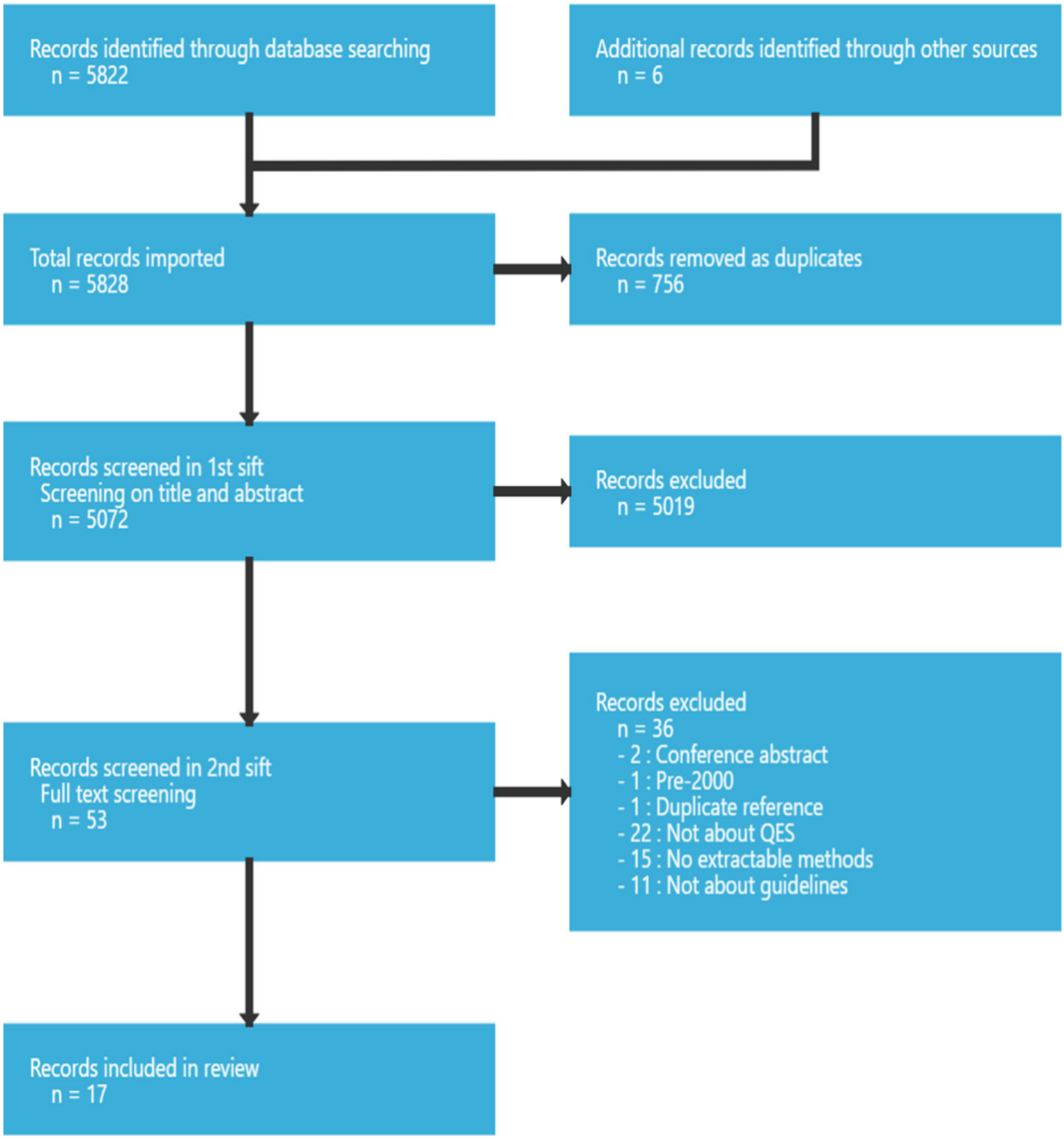

FIGURE 1 Study selection [Colour figure can be viewed at wileyonlinelibrary.com]

were included in the review. See Figure 1 for a PRISMA diagram summarizing the flow of papers.

3.2 | Study characteristics

Full references of included papers are provided in File S5. Table 1 provides a brief summary of each paper.

3.3 | Synthesis of results

3.3.1 | Methods for review protocol development or scope generation

Seven of the 17 included papers discussed or referred to the need for a scoping process or a process of review protocol generation before the searching and literature identification phase of the development of a guideline. In terms of the scoping or review protocol development itself, the critical tasks for the scoping phase of a guideline are described as identifying the interventions, stakeholders and contexts relevant to the guideline questions. This can be time consuming and adequate time needs to be set aside for this part of the process[7]; the decisions reached in these discussions directly inform the question and content of the review protocol.

Overall, there is broad agreement that a review protocol or scoping process should use expert input (e.g., a guideline committee, or service user organization) to derive the review question and the criteria for that review (e.g., using a PICO or SPICE format). The protocol is therefore the first stage of the process unless prior searching is required for evidence to inform the scope or

TABLE 1 Brief details of included papers

Author	Type of paper	Summary
Booth 2016	Guidance	A report funded by the EU as part of a series on evaluating complex interventions ("INTEGRATE- HTA"). The guidance document sets out a framework to enable reviewers to choose between different QES methods depending on the question they are asking.
Carroll 2017	BMJ Analysis	The analysis focusses on the need for successful guidelines to reflect patient views and argues that qualitative evidence is a key way to do this. The paper is not primarily a detailed methodological paper but contains some extractable methodological detail.
Downe 2019	Research article	The first in a series of three papers that have been written by a group of methodologists working with the WHO on guidelines that integrated QES. The authors examine the use of QES in developing clinical and health systems guidelines.
Flemming 2019	Analysis	This paper presents an overview of the ways QES can be used to address complex interventions
Glenton 2016	Manual/ handbook	Chapter 15 of the WHO handbook for guideline development specifically about using evidence from qualitative research to develop WHO guidelines.
Glenton 2019	Research article	The third in a series of three papers describing the use of qualitative evidence syntheses (QES) to inform the development of clinical and health systems guidelines by a team of methodologists who have worked with WHO. The WHO is increasingly using evidence derived from QES to provide information on acceptability and feasibility in its guidelines.
Gould 2010	Methodological report	Gould describes qualitative work done to support the production of two social care guidelines by the UK National Institute for Health and Care Excellence (NICE)
Hansen 2011	Methodological report	This article focuses on qualitative research synthesis in eliciting patients' perspectives as part of the growing drive to include patient views in policy and HTA
Knaapen 2015	Toolkit (chapter)	Chapter 2 of a GIN toolkit on patient and public involvement in guidelines. It contains practical ideas about how to conduct a qualitative evidence synthesis as part of the guideline development process.
Kristensen 2007	Manual/ handbook	The 2007 updated edition of the *Health Technology Assessment Handbook* that was issued by the Danish National Board of Health in 2001 as part of the fulfilment of the National Strategy for HTA. Contains some general detail about QES and also a specific chapter on assessment and syntheses [sic] of qualitative studies (section 4.2).
Lewin 2018	Commentary	Argues that the development of more "robust" (transparent) methods and tools for QES has widened the opportunities for QES to be used to inform health guidelines (in the context of the WHO).
Lewin 2019	Research article	This is the second in a series of three papers written by methodologists working with the WHO that examines the use of QES in developing clinical and health system guidelines. It specifically discusses using qualitative findings as part of evidence to decision frameworks.
NICE Manual 2018	Manual	The process manual used by NICE to produce clinical guidelines. The NICE manual includes details of synthesis for all the types of evidence it uses, not just qualitative evidence.
Ring 2010	Guidance	Guidance from NHS Quality Improvement Scotland about the various methods of QES that could be used in HTA.
Ring 2011	Research article	The authors conducted a systematic search to identify QES and reflect on the methodological approach used.
Swedish Agency for Health Technology Assessment and Assessment of Social Services (SBU) 2016	Manual	Swedish agency for health technology assessment manual for evaluating and synthesizing qualitative material.
Tan 2009	Evidence utilization report	Describes the use of qualitative research as evidence in a national clinical guideline program (National Institute for Health and Clinical Excellence—NICE, UK)

protocol (e.g., a QES). Scopes or protocols for QES may include a reflexivity statement.

The review protocol or scope should be made publicly available before the review commences in the same way as a quantitative systematic review would be registered.[21]

3.3.2 | Methods for identifying literature

Two thirds of the included papers contained information regarding the optimal methods of identifying evidence for a qualitative evidence synthesis. Most agreed that as part of the process there needed to be a systematic search of databases and pointed out that in many ways this was similar to quantitative database searching.

This focus on systematic searching represents a step-change from earlier in the 2000s when searching seems to have been less developed. However, some authors argue for systematic searching but also note that it may not be important to identify every available study, citing theoretical saturation as a possible endpoint.[6,22] QES conducted alongside a quantitative systematic review will be more likely to have more explicit inclusion criteria than a synthesis of qualitative studies that aims for theoretical saturation, where there might be a more iterative approach to searching and screening.[6]

Purposive sampling is also suggested. It is described as an iterative process of searching and screening, with the process being complete when the reviewers achieve "theoretic saturation" [sic] or "conceptual robustness."[9] It is frequently used as an adjunct to, or occasionally as a replacement for, exhaustive systematic database searching.

As a result of concerns over missing data, most authors recommend some kind of additional search method. Methods for additional searches that were mentioned include footnote and reference list checking, hand-searching key journals relating to the topic of the QES, forward citation searching (searching for relevant work by locating studies that cite earlier key studies), and author searching (searching for all publications by the author of a relevant work).

In summary, while most authors advocate some kind of systematic searching process, the complexities of identifying qualitative literature have led to approaches that try to reduce the volume of literature found by comprehensive searches (e.g., by using filters or more specific search terms), while also trying to mitigate the potential loss of relevant papers by adding in supplementary search techniques such as citation searching.

3.3.3 | Methods of study selection

Only three of the 17 papers gave any detail about considerations required in selecting qualitative studies. Two at some length,[23,24] and the other[21] in a single paragraph. Fundamentally the authors agreed that the process for study selection of qualitative literature should mirror the process that would be expected in a quantitative systematic review, with multiple reviewers comparing the paper with pre-specified inclusion and exclusion criteria, with any disagreements resolved by discussion or by the use of a third reviewer. This was especially the case when QES was being conducted alongside a quantitative systematic review.[24] None of the other papers discussed methods of study selection.

Study selections should be transparently reported, for example, through using a PRISMA diagram (Preferred Reporting Items for Systematic Reviews and Meta-analysis)[25] to show flow of studies through various stages of selection.[23,24]

There is a possibility of retrieving a large number of studies, especially through systematic searching, and it is recommended that in these cases reviewers select a sample.[23] It is difficult to quantify what constitutes a "large number" as it will, to a certain extent, depend on the emerging themes and concepts as well as on resources available and the timeframe of the review. Reviewers also need to be aware of introducing reviewer bias.[23]

3.3.4 | Methods of quality appraisal

The 10 included studies that discuss critical appraisal are broadly in favor of assessing the methodological quality of the studies included in a QES. This position is likely to be strengthened by the introduction and widening use of the GRADE CERQual tool[15] for assessing confidence in findings from QES since CERQual relies on a methodological assessment (among other things) of the studies included in the review findings.[9]

The latest edition of the NICE manual[22] states unequivocally that "Critical appraisal of qualitative evidence should be based on the criteria from the Critical Appraisal Skills Programme" (p. 106) and this is echoed by both in the WHO handbook[4] and the SBU methods manual.[21] More broadly, authors agree that studies should have some form of quality appraisal, preferably using one of the recognized appraisal systems for qualitative research.[23] The same advice is found in relation specifically to HTA.[13] Other authors, however, are more cautious and refer to lack of agreement about the value of critical appraisal of qualitative studies.[6,24] The GIN toolkit[9] succinctly summarizes the issue—"The use of

standardized "checklist" approaches has been strongly critiqued by some commentators, questioning how quality criteria modelled on the principles of positivist science can be applied to non-positivist qualitative research" (p. 33). However, in spite of this they provide a summary of the strengths and limitations of a range of checklists, including the CASP tool,[26] the Cochrane manual (Chapter 20),[27] the cabinet office Quality in Qualitative Evaluation tool,[28] and the Joanna Briggs Institute tool.[29]

It is also noteworthy that some specific approaches to QES (e.g., framework, meta-narrative and thematic synthesis) all have their own approaches to critical appraisal, whereas in other approaches such as meta-ethnography or grounded theory, critical appraisal is less important.[24] The overall consensus is that some form of critical appraisal should be conducted that appraises the methodological conduct of the study. Methods for assessing the content and validity of data are discussed in section 3.3.8.

3.3.5 | Methods of synthesis: General approaches

Much of the description of methods for the synthesis of findings from primary qualitative studies was presented as methods for specific QES methodologies. These specific approaches will be discussed after an outline of the generic methods referred to by other authors.

The Swedish HTA handbook for evaluation of qualitative studies[21] describes the evidence synthesis process as having four discrete stages. First, papers are read to give an overview of themes, then the papers are re-read and coded. No detail is given on the method of coding, but the manner in which it is described implies a process of emergent coding where codes are allowed to emerge from the included papers. As a second stage these "first level themes" are "distilled to form the second level theme." This appears to be an aggregative coding process of drawing together similar codes. Third, an interpretive coding phase is performed, described by the paper as follows—"Related second level themes are finally synthesised to an overall third level theme. Important patterns and associations among the second level themes are interpreted and problematised. The process is repeated until third level themes are set." The final stage in their synthesis process is "a general assessment of the scientific basis is made. Thereafter evidence graded results and conclusions are formulated," which appears to describe some assessment of review findings, like GRADE or GRADE CERQual.

While there is agreement that findings should be "added up" or compared and contrasted, the process of doing so inevitably masks their variability. This makes it easy to lose sight of the individuality of participants and their context that are the very heart of qualitative research. Two included papers highlighted how reviewers should strive to avoid this[9] and need to find the balance between splitting themes emerging from synthesis to the point that they are no longer useful, or lumping data together into themes that oversimplify or lose variation in the data.[23]

This need for descriptive or interpretive themes is driven by the nature and purpose of the QES, which is in turn dependent on the nature of the guideline being developed.[30] The outputs from framework syntheses or thematic syntheses are often as simple as a list of themes identified across included studies, with little or no interpretation, that can be used to "detail the needs, values, perceptions, behaviours and experiences of stakeholders within the guideline".[23] On a related note, it is suggested that QES used in guidelines tends to focus either on people's views about the interventions under scrutiny by the guideline, or it relates more widely to people's views and experiences of the condition underlying/addressed by the intervention(s) that the guideline is examining.[7] If interpretive findings are being produced then there is a need for transparency on the part of the reviewers to ensure that the interpretations are plausible and to show how they were arrived at.[23]

The use of Evidence to Decision frameworks (EtD) can be a driver for the style of the QES[23]:

The main purpose of an EtD-orientated QES is to generate a series of findings from the included data, which are directly focused on interventions addressed in the guideline, assessed for confidence and tailored toward acceptability, feasibility and equity, and the values that stakeholders attribute to the outcomes associated with the intervention. The findings are then added to the guideline EtD frameworks, prior to guideline panel consideration, (p. 8).

Infographics and logic models can also be incorporated into EtD frameworks in cases where the synthesis is intended to be explanatory or theory building.[7]

Overall, discussion about the general methods to be used for synthesis focuses on the continua between aggregative or integrative coding and more interpretive coding, and between the lumping and splitting of themes. This depends to some extent on the methodology used for the synthesis described in the next section.

3.3.6 | Methods of synthesis: Specific methodologies

Several authors provide brief (or occasionally in-depth) descriptions of methods of synthesis that can be used. It

is not the remit of this paper to reproduce general methodological detail about the various methods, but where authors have made comment on what makes a method suitable or unsuitable to produce a QES for a guideline development process, that has been included here.

There is a range of different QES methodologies available, some more developed than others. They predominantly reflect methods of primary qualitative research. The different methodologies sit broadly on a continuum between aggregative (or integrative) approaches that summarize themes and interpretive approaches that generate new interpretations of the data.[30] The Cochrane Qualitative and Implementation Methods Group recommend that the method of synthesis should only be chosen after the evidence is known and caution against prespecifying a methodology.[31] Epistemology is particularly important for some types of synthesis, with commentators arguing that the method of synthesis should be compatible with the epistemology of the included studies. Other methods may rely less on epistemology—for example best fit framework synthesis, narrative synthesis and thematic synthesis. In health services research and technology assessment a more pragmatic approach is taken with it being common to integrate different types of study within a single synthesis.[31]

Selection of an appropriate method is seen by authors as complex and dependent on many factors, especially the distinction between aggregative methods (where themes are integrated/aggregated) and interpretive methods (where the researchers try to add additional layers of interpretation over the data). The philosophical view of the researcher and purpose of the review can be driving factors,[6] as can the need for "pragmatic and relatively rapid methods of qualitative evidence synthesis" that might fit better with guideline developers' timelines.[8] Framework, narrative, and thematic synthesis are highlighted as particularly useful for answering questions about the uptake of interventions and for integrating quantitative and qualitative findings. This may make them particularly useful for developers of clinical guidelines - NICE already use some form of thematic synthesis in some of their guidelines.[8]

This identification of thematic synthesis methods as highly appropriate to guideline development is in line with the latest iteration of the NICE methods manual[22] which continues to identify thematic analysis as an appropriate methodology for analysing qualitative data. It advocates extracting "first level themes" into evidence tables (Evidence tables are detailed summaries of the content of each study included in a review or synthesis. These are normally incorporated into an appendix of the review or synthesis.). These evidence tables are then used to generate "second level themes" in the body of the synthesis. The manual also goes on to discuss (in passing) conceptual mapping, grounded theory, meta-ethnography and meta-synthesis, but notes that expertise in their use is needed (p. 107).

Most of the studies that specify methodologies refer predominantly to the same pool of synthesis methods: narrative synthesis,[12] meta-synthesis,[9,13] "imported concepts,"[13] meta-ethnography,[9,13] meta-study,[13] qualitative meta-summary[9,13] and framework analysis.[9] The older Danish HTA manual[12] mentions only meta-ethnography and narrative synthesis. This is likely because limited QES methodologies were available at that time. In a 2011 survey of 107 different QES, reviews using critical interpretive synthesis, meta-interpretation, qualitative cross-case analysis and grounded theory synthesis were found infrequently, and therefore their usefulness as methods of QES for HTA is unknown.[6]

By far the most comprehensive and well-developed guidance for selecting an appropriate QES method for HTAs is a report for the INTEGRATE-HTA project.[31] This project develops various criteria for QES and matches them to 19 different methodologies. Reviewers can select an appropriate method by aligning their needs with the various methods for conducting QES as outlined in the guidance's comprehensive tables that clarify a diverse range of considerations for each method. The project was directed specifically at HTA methods, but there seems no reason to suppose they would not be equally applicable to broader health guidelines.

Overall, thematic synthesis is the most frequently mentioned form of QES in guidelines and seems to be the most commonly used, with meta-ethnography and framework or best-fit framework synthesis as alternatives. This is primarily because these are the methods that are easier to use, and other methods have not been well tested, so they may be useful or not.[6] While a broad range of QES methodologies can be useful, the art is in selecting the appropriate methodology for the research question and research context.[31] The method of synthesis should only be chosen after the evidence is known.[30]

3.3.7 | Reporting standards

Reporting standards for QES were not discussed at great length in any of the included papers, possibly because different organizations have well established reporting standards internally. However, "...generic qualitative evidence synthesis reporting guidelines exist, others are being developed for particular methods, and standards are evolving to establish the level of confidence users can ascribe to the findings of such syntheses."[8]

There is a trend toward systematic review-like transparency in QES.[23,30] Historically, transparency has not been handled well by people reporting QES but work has

been undertaken to develop reporting standards for QES, such as the ENTREQ tool[32] and the eMERGe tool for meta-ethnography.[33] Newer tools have also emerged, notably RAMESES for realist synthesis.[34]

A useful minimum reporting standard has been used in work with the WHO.[23] The standard closely matches the reporting standards for quantitative systematic reviews (Cochrane reviews particularly) and suggests the characteristics and critical appraisal of each study should be presented in some detail, accompanied by a summary of themes (summary of qualitative findings) along with the confidence in those review findings and reasons for any downgrading. It also suggests providing a list of excluded studies, along with reasons for exclusion.

The guideline handbooks also briefly recommend approaches to reporting, with the WHO handbook[5] recommending the use of a summary of qualitative findings table that includes CERQual assessment (if there is one). SBU[21] also recommends the use of tabulation and the use of illustrative quotes where possible. The NICE manual[22] is more prescriptive and requires researchers to provide extensive evidence tables for all included studies containing the key information about the study. When CERQual is not being used, the NICE manual requires the production of evidence statements that summarize the evidence, its context and quality, and the consistency of key findings and themes across studies (meta-themes).

3.3.8 | Moving from evidence to recommendations

Papers discussed various aspects of the process of evidence-based recommendation making that fall generally into four categories—certainty in findings (including CERQual); frameworks (including evidence-to-decision frameworks); committees; and making recommendations from the evidence.

Certainty in findings from QES

Many guideline development agencies, including WHO and NICE require information on the confidence of findings that are used to underpin recommendations.[7,22] Since its publication in 2015, use of GRADE-CERQual has become the most common tool used as a summary measure when evaluating qualitative evidence for guideline development. A series of papers based on a WHO guideline were written by members of the original team who authored and devised the CERQual system for assessing the certainty in findings of qualitative evidence, and in the WHO papers they recommend the use of CERQual in guideline development.[4,7,23] The earliest mention of CERQual in the included papers is in the WHO guideline handbook qualitative chapter,[5] which contains a brief description of the components of CERQual as a tool to measure the level of confidence, in each of the findings of the QES. It also notes its similarity to GRADE for quantitative studies.

NICE recommend the use of CERQual somewhat more robustly. The manual notes that unless the qualitative evidence is very sparse or disparate (in which case a narrative approach is appropriate), the results of QES should be presented as summaries ("at outcome level") and should be assessed with GRADE CERQual. They present "evidence statements" (narrative summaries) as a less preferred alternative.[22]

Frameworks

The three papers in the recent WHO series[4,7,23] discuss at some length the use of EtD, an approach developed by the GRADE working group to increase transparency in moving from evidence and contextual considerations to implementable recommendations.[17,18] These EtD frameworks take the form of tables that draw together the key information necessary for guideline committees to make recommendations, including the PICO for the research question, summaries of the evidence, details of equality issues, feasibility issues, implementation consideration, and so on. They contribute to the overall transparency of the movement from evidence through discussion by a guideline committee or similar into recommendations but are not specific to QES. The evidence from QES can be added to the evidence section of the EtD framework alongside any quantitative evidence, along with its CERQual assessment,[6] Qualitative evidence does not always fit well within the "summary-based and compartmentalised structure" of the EtD framework. There are also implementation issues related to clinical guidelines and it is possible that evidence from QES that does not make it into the evidence section of the EtD framework can often be rewritten a little and turned into an implementation consideration.[4]

The only paper outside of the three WHO papers that discussed EtD frameworks stated that "[a] QES can be conducted separately or can be integrated with some form of quantitative synthesis. Within a guideline development process, findings from a QES will often be integrated with evidence of effectiveness in an evidence to decision (EtD) framework, used to formulate recommendations."[30] In another paper, a broader discussion about frameworks generally, the usefulness of frameworks in organizing data for a QES, and also for identifying gaps in qualitative data is highlighted.[4]

Committees

There was little discussion of the role of any kind of guideline or oversight committee in interpreting the

TABLE 2 Areas of agreement and opportunities for further development

Areas of agreement	Gaps/assumptions	Development opportunities
Protocol development		
• Review protocols are important, using SPICE, PerSPEecTiF or SPIDER[38] rather than PICO formats • Beneficial to involve lay people and experts in protocol development, but this is resource intensive and time consuming • Different frameworks for formulating research questions/protocols are developing, for example recent work on the PerSPEcTiF framework[41]	• The role of tertiary reviews ("reviews of reviews") is established in the quantitative literature (for example, the Cochrane manual,[27] chapter 22), and is often used for scoping reviews or mapping reviews to provide an overview of the field. There is no discussion of tertiary QES. • There is an unspoken assumption that a single QES would underpin an entire guideline	• What is the value (if any) of syntheses of existing QES? • How useful is a single generic QES for a guideline compared to specific QES for different elements of the guideline.
Identifying literature		
• structured searching using validated qualitative filters • some kind of supplementary searching is also common—reference list searching, citation searching, asking experts, or trawling gray literature • some support for introducing a concept of theoretical saturation to prevent searching becoming too onerous	Relevant data may be included in studies that are not directly relevant to the research question at hand	What is the optimum balance between inclusive searching and specific searching and sifting to identify relevant themes for QES?
Study selection		
• process mirrors quantitative—studies are matched against inclusion and exclusion criteria in the review protocol, usually in a two-stage process, firstly based on title and abstract, and then for papers that are not obviously excludable, at full text.	Assumes that standardized pre-specified protocol methods are superior to iterative, emergent qualitative methods.	Can an iterative approach to study selection be transparent enough to meet the transparency requirements for health guidelines?
Quality appraisal		
• researchers undertaking QES for the purposes of guideline development are more often in favor of transparent and systematic methods than researchers with a more pluralist approach to QES • agreement about the importance of using some kind of transparent process for quality appraisal is greater within the QES for guidelines literature reviewed for this study than among the general QES literature • CASP most commonly used tool though over 100 tools in circulation	• little agreement about the best way to measure the quality of a qualitative study because of methodological variation and differing views about what a "good" quality study looks like	• CAMELOT critical appraisal tool in development • Should studies of very low quality be excluded from any analysis or can they contribute to the overall analysis?
Synthesizing findings		
• the most common (and probably the most accessible) form of synthesis is a thematic analysis that goes through a process of aggregative coding, sometimes followed by interpretive coding	• It is uncertain whether some methods are more appropriate than others for particular types of question • most methods of QES have not been used often in guideline processes, and therefore it is unclear whether they are useful	Continuing evaluations of different QES methods used to underpin health guidelines.

TABLE 2 (Continued)

Areas of agreement	Gaps/assumptions	Development opportunities
Reporting		
• Lack of detailed discussion of reporting standards • The ENTREQ tool is commonly used as is some adaptation of the PRISMA standards for quantitative reviews. The Cochrane Effective Practice and Organisation of Care (EPOC) group have recently released a template for QES[42]	• Different organizations have their own in-house reporting standards. It is uncertain to what extent these overlap.	• How well do different reporting standards from major QES producers differ, and how can the differences be resolved?
Recommendations		
• GRADE and GRADE-CERQual have been a valuable addition to the decision-making process, but there is often no obvious link between the review findings (and any other information considered by the guideline producing bodies as evidence) and recommendations made by the committee (and the relative strength of the recommendation) • only WHO seem committed to using EtD frameworks	• It remains unclear how guideline committees/panels move from QES findings to making recommendations • lack of understanding about the processes by which committees use evidence to generate recommendations – no clear insight into committee use of different types of evidence (quantitative, qualitative) to approach different types of questions (effectiveness, feasibility, etc.)	How do guideline producing committees engage with different types of evidence (alongside their own beliefs, knowledge and experience), and how do they use that evidence to form recommendations?

evidence generated by QES and using it to develop recommendations, beyond the discussion reported above in relation to EtD frameworks. One example described is a process undertaken in the production of a social care focussed guideline (on dementia) where the evidence was searched for and reviewed by an academic review team, but the "weighting and synthesis" of evidence was done jointly with a guideline committee that included patients and carers.[35]

During the process of guideline production in committees, members may need to be reminded of relevant qualitative evidence, and this "champion" role might more easily be taken up by the producer of the synthesis, the methodologist or patient representatives.[9]

Making recommendations from the evidence

The process of making recommendations using the results of QES was not discussed in depth in any of the papers, with those that mention it mostly reporting that it is difficult to capture by simple steps and rules.[9] Normally, committees (in whatever form they take) make recommendations based on one or more systematic reviews, including any QES, alongside any other information that the committee consider to constitute "evidence" (e.g., EtD frameworks). However, this is not always the case and points out that sometimes confidence in QES or other types of evidence based on published studies may be overridden, for example, by human rights considerations or other overarching principles or normative values.[4]

So, although the evidence is primary, it is not the only consideration for guideline committees, and that is true of both quantitative systematic reviews and QES. The amount that any kind of evidence drives a decision about a particular recommendation should depend on the question being considered, and the judgements made should be supported by clear and transparent justifications.[6] There is little contained in the included papers to explain or clarify the process of using QES to inform recommendations.

4 | DISCUSSION

Overall, the literature relating to the use of QES within the context of guideline development seems to mirror large parts of the more general literature on QES, and this is of little surprise since the key people driving the development of QES methods in health and social care are also often the same people who are driving the agenda for using QES in guideline development processes.

In a world where evidence-based healthcare is dominated by the systematic review of randomized controlled trials (RCTs) by organizations such as NICE, Cochrane, WHO,

and so on, a model of QES that matches their already existing methods of standardization is likely to be more acceptable to them (and to fit better with their existing methods of interrogating evidence) and standardization of QES methods may be, in part, a strategic move by their advocates to make the methods more acceptable to organizations that have traditionally been sceptical of qualitative research. The Cochrane qualitative and implementation methods group (QIMP) have been instrumental in increasing the standardization of QES, most notably by supporting the development of GRADE—CERQual for assessing the level of confidence in summary qualitative findings in a way that clearly (and purposefully) matches the process of using the GRADE tool on quantitative pooled outcomes, and also by supporting methods of QES that can be integrated into or presented alongside Cochrane systematic reviews.

There are some sections of the review process where much of the literature was silent, for example, only three papers discussed study selection. It is unclear whether this was because they regarded it as less important, but none advocated a more traditional, emergent qualitative approach.

Overall, in this included literature, there is a level of excitement about the possibilities of using syntheses of qualitative evidence alongside traditional quantitative evidence, leading authors to cautiously proclaim a "new era" for qualitative research, supported by recent developments in QES methodologies, such as standardized methods for synthesis and for assessing the confidence that can be placed in the findings.[36]

In spite of this enthusiasm, there is still not universal agreement that qualitative evidence can be synthesized in a way that is meaningful or useful, or that standardization of methods of QES to make them more acceptable to the evidence based medicine movement is the best way to synthesize qualitative evidence. None of the included studies commented on the usefulness of more traditional reviews of qualitative evidence in developing guidelines. In large part, disagreements stem from the fact that primary qualitative evidence makes no claim to be generalizable, yet for a QES to be useful to a guideline producing committee, the committee needs to be able to argue that the evidence speaks to common experience. Although these arguments seem to be broadly ignored by researchers producing QES currently, the early days of QES were dogged by these arguments[37] since it "has emerged from the confluence of conventional systematic review methods with methods for primary qualitative research. With such a mixed heritage, and the juxtaposition of quite different epistemological positions, it is inevitable that the resultant tensions have generated considerable creative energy and significant methodological frictions."[38]

Once past the epistemological arguments about the philosophical feasibility of QES, the practical methods of performing a QES seem to have converged, in terms of their applicability to guidelines at least, over the past decade. There seems to be broad agreement over most stages of producing a QES to inform a guideline, even if the fine detail is not always consistent. Table 2 highlights similarities and areas of agreement in the reviewed literature as well as some differences and gaps where further investigation could be fruitful.

4.1 | Gaps highlighted by the review findings

This review has highlighted some areas where there appears to be enough consistency between different approaches to give a reasonable level of confidence in them. In some cases there seems to be better agreement between authors writing about QES methods for guideline production than for QES more generally, for example, in terms of the level of agreement around searching and around the need for critical appraisal of qualitative studies, both of which are contentious in the broader field of QES.[39]

There are also areas where there is less clarity, some of which are likely to be quite specific to QES in the context of guideline production, and therefore could probably not be resolved by wider searching of the literature.

As more and more QES are published in health and medicine, it becomes more likely that reviewers will find existing QES that wholly or partially answer their research questions. There is no discussion in the literature to explore how these may be used. Parallels in quantitative systemic reviewing include updating and using pre-existing reviews as evidence for committees; other reviewers use the inclusion lists from systematic reviews as a check that they have identified the relevant literature. In areas where several very similar systematic reviews exist a "review of reviews" or tertiary review can be conducted. None of those things is reported in the papers included here, but similar methods might be possible for QES.

Much of the literature included in this review contains the unspoken assumption that one guideline will require one QES. Only one included paper moots the possibility of multiple QES for one guideline,[7] however it is easy to imagine a guideline that contained questions that could be informed by several QES. There is no discussion in the included papers of how this might work in practice.

In terms of producing QES that are useful to guideline committees as part of the evidence-base they consider, is a standardized methodology best? Or is methodological pluralism more useful where the methodology can be a more pragmatic choice and take into account the time, resource and outputs that are wanted?

There are some 30 methods for conducting a QES,[31] and at least a further 10 methods are in development.[40] Few of those methods have been used frequently in producing QES for guideline development and so their utility is uncertain.

A related gap in this review is an understanding of the most useful way to provide guideline committees with the outputs of various reviews, both qualitative and quantitative alongside other kinds of evidence. Traditionally, guideline producers have prioritized reviews of RCT evidence to provide evidence of clinical efficacy, but the hierarchy of evidence for other types of outcome is less established. As QES become more standardized in their methods and transparency, can they become the principal evidence for certain types of guideline question? One series of papers discusses the use of EtD frameworks in some depth in one of the papers,[7] but these have not been the subject of robust evaluation and it is unknown how useful they are to committees.

Finally, there is little research, and none in this review, that explores how committees move from QES findings (or indeed quantitative systematic review findings) and the other information they are given as evidence—systematic reviews, expert testimony, real-world data, and so on, in the context of their own expertise and experience—to making decisions that produce guideline recommendations, and this is a fundamental question for guideline producers.

4.2 | Limitations

The content of this review was specifically limited to papers describing methods for using QES in guideline production. As a result, it does not cover the large, and growing, corpus of literature dealing with the topic of QES generally. There are good published overviews of the development of QES, notably through leadership from the Cochrane Qualitative and Implementation Methods Group.[16] Fortunately, several of the key authors writing about the use of QES in guideline development are also key authors in the field of QES in health more generally, so to a large extent the literatures inform and reflect each other.

For the purposes of time and resource, this review did not consider the implications of the growth in methods for integrating qualitative and quantitative data to produce mixed-methods reviews. As technologies for producing standardized QES and standardized systematic reviews develop, researchers are becoming interested in integrating qualitative and quantitative data in the aspiration that the whole may be greater than the sum of the parts.

5 | CONCLUSIONS

The use of QES to inform the production of health guidelines is growing as the methods for producing them become more clearly defined and more standardized. Methods for producing QES for guideline committees tend to be similar to quantitative systematic review methods in terms of searching, appraisal of evidence, systematic management of data and presentation of results. While this allows greater transparency and greater accountability, it could be argued that it is less "true" to the principles of being led by the data, which are fundamental to most qualitative research.

Recent developments in QES mean that there is broad agreement about how QES can be produced to help inform guidelines, but further research is needed to establish whether guideline-producing committees find QES useful to their deliberations, whether they could be done or presented differently to make them more useful and, perhaps most importantly, how committees use QES to inform their decision making alongside traditional systematic reviews of effectiveness.

CONFLICT OF INTEREST

Chris Carmona is a part time employee of the UK National Institute for Health and Care Excellence.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the supplementary material of this article.

ORCID

Chris Carmona https://orcid.org/0000-0001-9855-3575

REFERENCES

1. World Health Organization. WHO guidelines approved by the Guidelines Review Committee. https://www.who.int/publications/guidelines/atoz/en/. Accessed July 19, 2020.
2. National Institute for Health and Care Excellence (NICE). Guidance and advice list. https://www.nice.org.uk/guidance/published?type=apg,csg,cg,cov,mpg,ph,sg,sc,dg,hst,ipg,mtg,qs,ta. Accessed July 19, 2020.
3. Langlois EV, Tunçalp Ö, Norris SL, Askew I, Ghaffar A. Qualitative evidence to improve guidelines and health decision-making. *Bull World Health Organ*. 2018;96(2):79-79A.
4. Glenton C, Lewin S, Lawrie TA, et al. Qualitative evidence synthesis (QES) for guidelines: paper 3 - using qualitative evidence syntheses to develop implementation considerations and inform implementation processes. *Health Res Policy Syst*. 2019;17(1):74.
5. Glenton C, Lewin S, Norris SL. *World Health Organization Handbook for Guideline Development*. 2nd ed. Geneva: WHO; 2016.
6. Ring N, Jepson R, Ritchie K. Methods of synthesizing qualitative research studies for health technology assessment. *Int J Technol Assess Health Care*. 2011;27(4):384-390.

7. Lewin S, Glenton C, Lawrie TA, et al. Qualitative evidence synthesis (QES) for guidelines: paper 2 - using qualitative evidence synthesis findings to inform evidence-to-decision frameworks and recommendations. *Health Res Policy Syst.* 2019;17(1):75.
8. Carroll C. Qualitative evidence synthesis to improve implementation of clinical guidelines. *BMJ (Clinical Research Ed)*. 2017;j80:356.
9. Knaapen L, Colvin CJ, Cowl J, van der Weijden T. How to include qualitative research on patient views in guidelines. *GIN Public Toolkit*. 2015; 28-40.
10. Tan TPY, Stokes T, Shaw EJ. Use of qualitative research as evidence in the clinical guideline program of the National Institute for Health and Clinical Excellence. *Int J Evid Based Healthc*. 2009;7:169-172.
11. Sandelowski M, Barroso J. *Handbook for synthesizing qualitative research*. New York: Springer Publishing Company; 2006.
12. Kristensen FB, Sigmund H, eds. *Health Technology Assessment Handbook. Danish Centre for Health Technology Assessment*. Copenhagen: National Board of Health; 2007.
13. Hansen HP, Draborg E, Kristensen FB. Exploring qualitative research synthesis: the role of patients' perspectives in health policy design and decision making. *Patient*. 2011;4 (3):143-152.
14. Schünemann HJ, Brożek J, Guyatt GH, Oxman A, eds. *GRADE Handbook for Grading Quality of Evidence and Strength of recommendations*. The GRADE Working Group; Available from guidelinedevelopment.org/handbook. 2013.
15. GRADE CERQual. Confidence in the Evidence from Reviews of Qualitative Research. https://www.cerqual.org/. Accessed September 14, 2019.
16. Cochrane Qualitative and Implementation Methods Group. https://methods.cochrane.org/qi/welcome. Accessed October 14, 2019.
17. Alonso-Coello P, Schünemann HJ, Moberg J, et al. GRADE evidence to decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: introduction. *BMJ*. 2016;353:i2016.
18. Alonso-Coello P, Oxman AD, Moberg J, et al. GRADE evidence to decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: clinical practice guidelines. *BMJ*. 2016;353:i2089.
19. *NVivo Qualitative Data Analysis Software* [computer program]. Version 122018.
20. EPPI *Reviewer* [computer program]. Version 52018.
21. Swedish Agency for Health Technology Assessment and Assessment of Social Services (SBU). *Evaluation and Synthesis of Studies Using Qualitative Methods of Analysis*. Stockholm: Swedish Agency for Health Technology Assessment and Assessment of Social Services (SBU; 2016.
22. National Institute for Health and Care Excellence (NICE). *Developing NICE guidelines: the manual*. Manchester: NICE; 2018.
23. Downe S, Finlayson KW, Lawrie TA, et al. Qualitative evidence synthesis (QES) for guidelines: paper 1 - using qualitative evidence synthesis to inform guideline scope and develop qualitative findings statements. *Health Res Policy Syst.* 2019;17(1):76.
24. Ring N, Ritchie K, Mandava L, Jepson R. *A Guide to Synthesising Qualitative Research for Researchers Undertaking Health Technology Assessments and Systematic Reviews*. Glasgow: Quality Improvement Scotland (NHS QIS); 2010.
25. Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA). www.prisma-statement.org. Accessed September 14, 2019.
26. Programme CAS. CASP Qualitative Checklist; 2019. https://casp-uk.net/casp-tools-checklists/ Published 2018. Accessed 14/09/2019.
27. Higgins J, Thomas J, Chandler J, et al. *Cochrane Handbook for Systematic Reviews of Interventions*. 6th ed. www.training.cochrane.org/handbook. Cochrane; 2019.
28. Spencer L, Ritchie J, Lewis J, Dillon L. Quality in Qualitive Evaluation: A framework for assessing research evidence; 2003. https://webarchive.nationalarchives.gov.uk/20140619055846/http:/www.civilservice.gov.uk/wp-content/uploads/2011/09/a_quality_framework_tcm6-38740.pdf. Accessed September 14, 2019.
29. Joanna Briggs Institute. JBI Critical Appraisal Checklist for Qualitative Research; 2017. https://joannabriggs.org/ebp/critical_appraisal_tools. Accessed September 14, 2019.
30. Flemming K, Booth A, Garside R, Tuncalp O, Noyes J. Qualitative evidence synthesis for complex interventions and guideline development: clarification of the purpose, designs and relevant methods. *BMJ Glob Health*. 2019;4:e000882.
31. Booth A, Noyes J, Flemming K, et al. Guidance on choosing qualitative evidence synthesis methods for use in health technology assessments of complex interventions. [Online]. In: Integrate - HTA; 2016.
32. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Med Res Methodol*. 2012;12(1):1-8.
33. France EF, Cunningham M, Ring N, et al. Improving reporting of meta-ethnography: the eMERGe reporting guidance. *BMC Med Res Methodol*. 2019;19(1):1-13.
34. The RAMESES Project. https://www.ramesesproject.org/. Accessed September 14, 2019.
35. Gould N. Integrating qualitative evidence in practice guideline development: meeting the challenge of evidence-based practice for social work. *Qualitative Social Work: Research and Practice*. 2010;9:93-109.
36. Lewin S, Glenton C. Are we entering a new era for qualitative research? Using qualitative evidence to support guidance and guideline development by the World Health Organization. *Int J Equity Health*. 2018;17:126.
37. Barnett-Page E, Thomas J. Methods for the synthesis of qualitative research: a critical review. *BMC Med Res Methodol*. 2009;9 (1):1-11.
38. Booth A. *Acknowledging a dual heritage for qualitative evidence synthesis: harnessing the qualitative research and systematic review research traditions*. Sheffield, England: University of Sheffield; 2013.
39. Carroll C, Booth A. Quality assessment of qualitative evidence for systematic review and synthesis: is it meaningful, and if so, how should it be performed? *Res Synth Methods*. 2015;6(2): 149-154.
40. Noyes J, Booth A, Cargo M, et al. Cochrane qualitative and implementation methods group guidance series-paper 1: introduction. *J Clin Epidemiol*. 2018;97:35-38.
41. Booth A, Noyes J, Flemming K, Moore G, Tunçalp Ö, Shakibazadeh E. Formulating questions to explore complex interventions within qualitative evidence synthesis. *BMJ Glob Health*. 2019;4(suppl 1):e001107.

42. Glenton C, Bohren M, Downe S, Paulsen E, Lewin S. EPOC Qualitative Evidence Synthesis: Protocol and review template; 2019.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Carmona C, Baxter S, Carroll C. Systematic review of the methodological literature for integrating qualitative evidence syntheses into health guideline development. *Res Syn Meth*. 2021;1–15. https://doi.org/10.1002/jrsm.1483