

Synthesizing Qualitative Research: Reflections and Lessons Learnt by Two New Reviewers

International Journal of Qualitative Methods
Volume 17: 1–14
© The Author(s) 2018
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1609406918768014
journals.sagepub.com/home/ijq


Dia Soilemezi¹  and Skaiste Linceviciute²

Abstract

Systematic synthesis of qualitative studies is widely used in health and social care. Regardless of the topic area, researchers need to consider several decisions when it comes to the planning and implementation of qualitative synthesis. As junior reviewers, we reflect on potential challenges and pitfalls in planning and conducting a synthesis of qualitative evidence. This article aims to elaborate on a number of key issues in order to provide insights and options on how to avoid or minimize these issues, especially for new reviewers and research students. This article examines difficulties in different stages and presents some examples of how intellectual and technical issues can be approached and resolved, including how to ensure effective identification of the relevant research to answer the review question? What are the potential pitfalls during the screening and evaluation process? The implications of different issues are examined and potential directions are discussed.

Keywords

systematic review, evidence synthesis, quality appraisal, qualitative methods, health policy

What Is Already Known?

There is a proliferation of methods for synthesizing qualitative research, and there are high-profile organizations providing systematic guidance on how to conduct and report a systematic review. This methodology is well established, and yet conducting a qualitative synthesis is often stressful and resource intensive. What is missing is a practical guide for how to not only navigate the decisions and process but also overcome technical obstacles that are bound to arise.

What This Paper Adds?

Following our reflections on the challenges of synthesizing qualitative evidence, this article presents a number of decisions and technical challenges and provides a practical guide for new reviewers to manage difficulties and work their way through the steps of a qualitative evidence synthesis. Careful planning involves being systematic in the methods, but also planning the management of the process, which is often underestimated.

Introduction

Synthesizing qualitative research has become a useful and popular tool to inform policy- and evidence-based health care in

recent years (Noyes, Popay, Pearson, Hannes, & Booth, 2008; Pope & Mays, 2006a). Systematic reviews can prove invaluable for busy practitioners as they combine results from many studies, provide up-to-date summarized evidence, and disseminate them in an unbiased and rigorous manner (Dixon-Woods, Agarwal, Young, Jones, & Sutton, 2004; Pope & Mays, 2006a). Topics and types of systematic reviews can vary depending on available evidence, resources (scoping vs. comprehensive), methodological viewpoints, and purpose. The advantage of systematic reviews is that they examine all the available literature and combine primary research studies related to a specific phenomenon or question to reveal a new explanation and deeper insights of the particular phenomenon that is not possible from a single study (Erwin, Brotherson, & Summers, 2011). Ultimately, systematic reviews aim to enhance our

¹ Department of Psychology, University of Portsmouth, Portsmouth, United Kingdom

² School of Health Sciences and Social Work, University of Portsmouth, Portsmouth, United Kingdom

Corresponding Author:

Dia Soilemezi, Department of Psychology, University of Portsmouth, King Henry Building, King Henry I Street, Portsmouth PO1 2DY, United Kingdom.
Email: dia.soilemezi@port.ac.uk



understanding and provide evidence in a way that allows transferability, to identify research gaps for further exploration, prevent unnecessary duplication of research, improve clinical outcomes for the patients, and guide evidence-based clinical decisions (Erwin et al., 2011; Pearson, 2004).

In synthesizing qualitative evidence, there is a proliferation of methods, with many approaches sharing common structures in their synthetic process or epistemological approach, but also strategic differences. For example, some approaches allow the generation of theories, such as meta-ethnography and grounded theory; some are solely used in qualitative research, whereas other approaches (such as thematic synthesis, realist synthesis, and critical interpretive synthesis (CIS) allow the integration of mixed methods design (Saini & Shlonsky, 2012). Some require the inclusion of similar study designs (e.g. grounded theory, meta-interpretation); others may include multiple study approaches (e.g. thematic synthesis, meta-ethnography, and meta-study) in their analysis (Barnett-Page & Thomas, 2009; Booth, 2016; Booth et al., 2016).

As the number of systematic reviews increases, so does the complexity and bewildering variety of choice around it. Conducting a systematic review (either on its own or as a part of a mixed method project) comes at a cost: It can be an extremely timely and resource intense activity (Kavanagh, Campbell, Harden, & Thomas, 2012). Specifically in qualitative evidence synthesis, the complexity of methods and the limited guidance can increase time and resource intensity. To assist reviewers with this laborious task, high-profile organizations provide systematic guidance (such as the Cochrane Collaboration, the Center for Reviews and Dissemination, the Campbell Collaboration, the Joanna Briggs Institute [JBI], the Systematic Review Data Repository, and the Evidence for Policy and Practice Information [EPPI] center) and there are published guidelines for reporting the synthesis of qualitative research (Tong, Flemming, McInnes, Oliver, & Craig, 2012). What is missing however is a practical guide about how to navigate the process and potential options and overcome technical obstacles that are bound to arise on the way. For example, the existing guidelines suggest the need to identify multiple databases when searching for relevant studies, but give little, if any, practical guidance on how many to include or how to use each database, handle downloads, and save the results.

As part of our own PhD degrees, we have each conducted a qualitative synthesis and encountered many challenges. Although standardized steps were followed, there were differences in how our syntheses were conducted and the important decisions we had to consider. This learning led us to write this article. By sharing our experiences and reflections, we aim to highlight challenges and technical difficulties and present some ideas on how to overcome them. Knowing some of the issues in advance can be very helpful in order to prepare the team's skills and resources without feeling overwhelmed or unclear on how to move forward. Our article is intended to help new reviewers and research students and it should be used in addition to existing guidance for conducting systematic reviews and choosing synthesis methods. The structure of this article follows the four

stages of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram (Table 1): primary study identification, screening, eligibility of studies, and inclusion and synthesis of findings.

Primary Study Identification

This phase includes formulating a review question, developing a protocol, identifying relevant research to answer your review question, and saving your search results.

Formulating a research question. Identifying a synthesis topic is essential for formulating the key question(s) a synthesis will address, and in most cases it involves a topic of intellectual interest to the reviewers. If a qualitative review already exists, the reviewers need to consider the value of conducting another similar review along with other issues (e.g., Is the existing review really systematic? Is it out of date? Does it answer the question in mind?). It may be very challenging for new reviewers to turn a synthesis topic into articulated, unambiguous, and precise key question(s) to develop a scientifically rigorous and pertinent review of evidence. A review question needs to explore an important and relevant issue to practitioners and/or patients under a certain context, looking at important outcomes, and ideally should be informed by patient and public involvement (Chalmers & Glasziou, 2009). The review question thus needs to be clear, justified, and focused with specific objectives (Rojon & Saunders, 2012). Reviewers are therefore advised to use the PICOD (population, the phenomena of interest, the context, the outcome of interest and design) mnemonic to construct a clear, specific and meaningful question(s) for the qualitative synthesis. Additional search strategy tools such as the PICOC (patient/population, intervention, comparison, outcomes, and context); CHIP (context, how the study was conducted, issues examined, and people involved); SPICE (setting, perspective, intervention/phenomenon of interest, comparison, and evaluation); ECLIPSE (expectations, client group, location, impact, professionals involved, and service); CIMO (context, intervention, mechanisms, and outcomes); and SPIDER (sample, phenomenon of interest, design, evaluation, research type) frameworks, among others, have been proposed as alternatives to use and complement the PICOD tool for qualitative evidence synthesis (Booth et al., 2016; Cooke, Smith, & Booth, 2012; Stern, Jordan, & McArthur, 2014). For a comprehensive list of mnemonics used for formulating qualitative questions, see also Booth (2016) and Booth et al., (2016).

This step needs adequate consideration to reflect the team's expertise, resources, and interests. The importance of this cannot be overstated. Focused and well-defined questions and objectives are more likely to identify appropriate and manageable citations to answer the review question at its core (Rojon & Saunders, 2012). If the review question is unclear, this may impact on the time and effort needed to complete the next phases, resulting in uncertainties at every step of the process. For example, the search strategy (which is heavily guided by

Table 1. Examples of Challenges and Potential Considerations When Synthesizing Qualitative Evidence.

Steps in Qualitative Evidence Synthesis (Guided by PRISMA Diagram)		Challenges to Consider	Possible Considerations
1. Primary study identification	1.1. Formulating a review question	<ul style="list-style-type: none"> • Is this a clear, focused, and answerable review question? • Is this an important question? • Is it a fixed or flexible question? 	<ul style="list-style-type: none"> • Use at least one search strategy mnemonic: PICOC, SPICE, ECLIPSE, CIMO, SPIDER, and so on • Review team should involve academic and clinical expertise and PPI • Consider the team's expertise, resources, and anticipate refinement of the review question
	1.2. Developing a review protocol	<ul style="list-style-type: none"> • Does the protocol clearly describe the plan of action? • How can the plan be described in emerging review questions? 	<ul style="list-style-type: none"> • Follow PRISMA-P checklist • Delay the publication of protocol until some uncertainties are resolved
	1.3. Identifying relevant research to answer the review question	<ul style="list-style-type: none"> • Which databases are right to use? • How to locate qualitative research? • Is exhaustive or purposive sampling more appropriate? • Is the search too wide/too narrow? • Are all databases available to use? • Will other language(s) and time frames need to be included? 	<ul style="list-style-type: none"> • Carry out scoping search in multidisciplinary database • Seek advice from librarians and information specialists to refine the search strategy and combine key terms with free-text words • Depends on review type and question. If driven by theoretical saturation, consider scoping search before purposive searching • If too wide, set limitation criteria; if too narrow, explode the key terms and consider including supplementary search methods • If not, explain any diversion from protocol • Decide if this is meaningful for the review question and consider the impact on forming and interpreting the analytical themes
	1.4. Saving the search results	<ul style="list-style-type: none"> • What is the best way to save imported results from different databases? • How to import the results? 	<ul style="list-style-type: none"> • Be aware of idiosyncrasies of each database and keep an audit trail of imports and duplications • If direct import to reference manager proves problematic, consider importing first on word and then transfer to software
2. Screening of identified studies	2.1. Screening stages	<ul style="list-style-type: none"> • Is screening titles open to missing qualitative studies? • What is the most efficient way to screen? • Is single screening biased? • Is screening clear to all reviewers? 	<ul style="list-style-type: none"> • If screening titles only, remain inclusive until the abstract stage • Consider screening titles and abstracts simultaneously • Consider full or proportional double screening • Pilot the screening and remain close to the protocol
3. Eligibility of studies and methodologies	3.1. Assessing full-text papers	<ul style="list-style-type: none"> • How to deal with unavailability of full-text papers? 	<ul style="list-style-type: none"> • Consider interlibrary loans, contact authors, and set deadline for response
	3.2. Appraising quality	<ul style="list-style-type: none"> • How to appraise the eligible studies? • What cutoff criteria to use? 	<ul style="list-style-type: none"> • Choose an existing checklist and expand if appropriate • Be mindful of cutoff criteria and judge individual papers on unique contribution, consider year and journal published, and contact authors
4. Inclusion, synthesis, and reporting	3.3. Excluding (or not) papers	<ul style="list-style-type: none"> • How to deal with low-quality studies? 	<ul style="list-style-type: none"> • Run a sensitivity analysis
	4.1. Extracting data	<ul style="list-style-type: none"> • How to tackle data extraction task? • What data to extract? 	<ul style="list-style-type: none"> • Ideally, two reviewers complete data extraction in tandem with quality appraisal • Use and adapt previous extraction forms
	4.2. Synthesizing the findings	<ul style="list-style-type: none"> • What analytic approach to choose? 	<ul style="list-style-type: none"> • It depends on the scope, available studies (number and type) and resources; follow guidelines and document the analytical steps

(continued)

Table 1. (continued)

Steps in Qualitative Evidence Synthesis (Guided by PRISMA Diagram)	Challenges to Consider	Possible Considerations
4.3. Writing up the findings	<ul style="list-style-type: none"> • What steps and procedures to report? • How to draw conclusion and recommendations for practice? 	<ul style="list-style-type: none"> • Follow ENTREQ checklist and PRISMA flowchart; consider using mind maps, charts, figures, and plain text • Plan how to integrate findings and report negative cases, involve PPI and clinicians to validate interpretations and recommendations

Note. PICOC = patient/population, intervention, comparison, outcomes, and context; SPICE = setting, perspective, intervention/phenomenon of interest, comparison, and evaluation; ECLIPSE = expectations, client group, location, impact, professionals involved, and service; CIMO = context, intervention, mechanisms, and outcomes; SPIDER = sample, phenomenon of interest, design, evaluation, and research type; PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses; PPI = patient and public involvement; ENTREQ = Enhancing transparency in reporting the synthesis of qualitative research.

the research question) may be ineffective, screening criteria will be unclear and so will the inclusion/exclusion decisions, and the data extraction will be troublesome. Our experiences indicated that when more time is spent in focusing on the review question, there would be less cost to the review team in time and confusion. However, if the review question is too narrow, then the disadvantage is that the review may not cover a phenomenon fully and the reviewers may end up with a limited number of studies that might not allow for a meaningful analysis to answer the review question. This may be the case in a fairly new topic area, on which not much research has been published or is ongoing. A scoping search in advance and expertise in the team on the topic area are thus valuable in guiding the formulation of a focused and answerable review question.

Depending on the review objectives and available data, it is also possible to have an emerging review question and refine it during the process (Booth et al., 2016; Rojon & Saunders, 2012). In this case, the reviewers need to be creative and comfortable with this highly iterative process as their research question serves the purpose of general direction and will be continuously modified in the review process. This approach, as any other, requires the documentation of the review process and decisions made at every stage to remain systematic and transparent.

Developing a review protocol. It is essential to develop a protocol, a detailed and transparent plan of action. It should specify a priori the rationale for the review, explaining the review question, the type of searches, studies and methodologies (e.g., purely qualitative or mixed methods design) to be included, the team involved, and time frame (or expected milestones) in as much detail as possible (Butler, Hall, & Copnell, 2016; Moher et al., 2015). The protocol typically provides a point of reference to reduce uncertainties in the team as well as to anticipate potential difficulties (e.g., which studies to include or exclude) in the synthesis process (Moher et al., 2015; Ring, Ritchie, Mandava, & Jepson, 2011). The review protocol needs to be published before the review commences to promote scientific transparency (identify accuracy or deviations), minimize potential bias, enable public access, and avoid duplication

(Moher et al., 2015; Pearson, 2004). Published protocols may also promote the trustworthiness of review outcomes and promote the credibility of findings (Butler et al., 2016). In qualitative evidence synthesis, bias may refer to researchers assumptions and input (Hannes, 2011) as well as decisions and alterations made throughout the review process that may impact on the way the eligible studies were identified (eligibility and selection of studies), analyzed (coded and synthesized), and reported at the end (Moher et al., 2015). Keeping an audit trail of decisions and disagreements, involving of at least two reviewers in each stage, and using specified and clear inclusion/exclusion criteria for each stage are some parameters that should be considered and documented in the protocol to help minimize bias.

This task may be particularly challenging for the reviewers when using an emerging question, which means drafting an iterative protocol that is subject to ongoing changes through the review process (Booth et al., 2016). As mentioned above, reviewers need to be aware and comfortable about the ongoing uncertainties and iterations of this approach as a component of the process and document all steps, decisions, and potential disagreements. Regular updates of the protocol are suggested as another way to remain transparent in this approach (Booth et al., 2016).

Review protocols can be published via registering with an organization (e.g., Cochrane, JBI) or in academic journals or on the international prospective register of systematic reviews (PROSPERO), an online-free database to register and regularly update systematic reviews on topics regarding health and social care from around the world. The PRISMA-P (protocols) checklist, although designed for quantitative reviews, can be useful for developing and reporting a robust protocol for qualitative synthesis (Moher et al., 2015). In some cases (e.g., in funded reviews), peer reviewing the review protocol may be required.

Identifying relevant research to answer your question. Once the review question is well-developed and focused, a search strategy is required. This will involve where to search (which and how many databases), with or without hand searching, for what time period, age-group(s), and in what language(s). This task may raise challenges as searching for qualitative evidence is

not as well developed as for quantitative evidence (Flemming & Briggs, 2007; Ring et al., 2011). Some experimentation or a scoping search is therefore invaluable to decide on the type and number of electronic bibliographic databases (including or excluding other resources) and thus refine the search strategy. This is typically dependent on the methodological approach (purposeful sampling vs. comprehensive searching), type of review (interpretive or aggregative), informed by previous scoping searches, and also dependent on the size of available literature for the given topic (Dixon-Woods et al., 2007; Gallicher et al., 2013; Pope & Mays, 2006b). For example, for a nursing-related topic, CINAHL is recommended to locate qualitative studies (Evans, 2002; Flemming & Briggs, 2007). At this stage, librarians and information specialists are crucial to the development of an efficient search strategy to target a specific domain to a high sensitivity, precision, and specificity (Jenkins, 2004).

The first step is to identify which databases to use to reflect the nature of the research question(s). It is important to search each database separately, bearing in mind that some may use different qualitative filters, symbols for truncation of terms (#, \$, *, !), or thesaurus terms (Booth, 2016; Shaw et al., 2004). Connectors (AND, OR, NOT) are the same in all databases. It is common to use “?” (e.g., institutional?ation) to account for U.S./UK spelling differences (Jenkins, 2004). It is also advisable to use multiple and specialist databases to ensure exhaustive searching and gather multidisciplinary literature if relevant (e.g., a combination of health related [MEDLINE, PsycINFO], social care [Social Sciences abstracts], and nursing [CINAHL] databases may be appropriate when exploring experiences of a health intervention). Exhaustive searching ensures that all relevant studies in the review topic could be captured to answer the review question adequately and thus minimizes the bias. However, in reviews that seek theoretical saturation, there is not a need for exhaustive searching and a more purposive sampling is considered the most suitable approach. However, this sampling has been criticized for being subjective, not reproducible, or systematic (Booth, 2016; Noyes et al., 2013). In this case, a scoping review is required to ensure inclusion of appropriate sample to identify papers with relevant characteristics and negative cases in order to provide a holistic interpretation of the review question.

Typically, the search strategy is guided by the review question and the search tool (e.g., PICOD, SPIDER), which helps to brainstorm relevant key concepts, context, and disciplines to be included as search terms. Depending on the review question, not all elements of PICOD are needed as search terms. However, combinations of these terms with free-text words (.tw.) as written by authors, synonyms, and other terms (e.g., using a methodological study filter, or broad qualitative methods search terms, such as findings, “interviews,” “qualitative,” “audio recording,” “grounded theory,” “thematic analysis,” etc.) are strongly recommended in order to maximize retrieval of relevant papers (Booth, 2016; Flemming & McInnes, 2012; Jones, 2004; Shaw et al., 2004; Wilczynski, Marks, & Haynes, 2007). For example, in one of our recent reviews (Soilemezi,

Drahota, Crossland, & Stores, 2017), the search strategy involved five sets of search terms (people with dementia, carers, professionals, home environment, and qualitative research) to capture the review question/title “The role of the home environment in dementia care.” Equally, using a plethora of key terms is important. Fewer key terms are likely to result in omission of references that might contribute important insight and thus might not guarantee answering the question effectively; broad (sensitive) searches might result in hundreds (even thousands) of irrelevant papers that will prolong the review. Getting the correct balance (sensitivity and specificity) in the search strategy is important. One option would be to check the initial search results for key studies (if known) and revise the strategy (e.g., add or remove terms) accordingly.

Some searching of index terms and testing of these (on one line so to amend easily if needed) can be undertaken before finalizing the search filters and final searches (Jenkins, 2004). Arguably, this process may be intellectually challenging and deeply frustrating, especially when searches return a large number of irrelevant references (false positives). This may happen because individual qualitative studies are context-specific and the location of qualitative research in electronic databases is both complex and challenging and often lacks appropriate index terms and abstracts (Booth, 2016; Evans, 2002; Flemming & Briggs, 2007; Pope & Mays, 2006b; Shaw et al., 2004). If the search results in a large number of citations, it is worth considering limiting the irrelevant hits by population age, (e.g., “adults” NOT “children”), publication type, and so on. On the other hand, if the search is not retrieving relevant results, the search can be expanded by exploding the key terms (exp) and by including titles and abstracts (.ti, ab.), as identifying relevant articles purely from the title can be difficult (Flemming & Briggs, 2007). In any case, it is important to report on which database this exercise was taken and be aware of possible limitations of transferring filters from one search interface to another (Jenkins, 2004). For example, the index terms used in one database may not be relevant to another.

Due to the poor bibliographic indexing of the qualitative research and despite explicit and comprehensive search strategies and combination of terms, it is possible that relevant studies may still be missed (Atkins et al., 2008; Evans, 2002; Noyes et al., 2008; Saini & Shlonsky, 2012). Identifying studies’ methods can be limited, depending on the database used (e.g., nursing and social work databases, such as CINAHL, use more qualitative indexed terms than medical databases), and sometimes due to the noninformative or descriptive nature of the qualitative titles and abstracts (Atkins et al., 2008; Evans, 2002; Flemming & Briggs, 2007; Ring et al., 2011; Saini & Shlonsky, 2012; Shaw et al., 2004). In our experience, another challenge at this stage was that institutional subscription access to chosen databases was not available for the duration of the review. As this stage is protocol-driven, any amendments made and reasons should be reported (Jenkins, 2004).

These limitations make it appropriate to include supplementary search methods and combine systematic database searching with supplementary searching: citation pearl

growing (Booth, 2016; Cooke et al., 2012), hand searching of important journals and/or citation lists (Britten & Pope, 2012; Jenkins, 2004; Jones, 2004; Ring et al., 2011), “snowballing” and contacting key authors, area experts (Gallacher et al., 2013; Noyes et al., 2013; Pope & Mays, 2006b), searching for books (Ring et al., 2011), and for gray literature (reports, thesis, not indexed in major databases) to minimize publication bias (Toews et al., 2017). Booth (2016) argued the importance of searching the references of the included full papers and suggested that all reviewers should include this “*simply as standard good practice*” (p. 16). In our experience, Google Scholar was helpful in retrieving books and gray literature. Also, for our reviews, contacting authors directly was partly successful (only a small number replied and some of them with a delay of few weeks); thus reviewers may need to balance the effort, time, and cost.

Choosing whether to limit your search to specific language(s) and years is another decision depending on the skills and resources available (e.g., time, funding for translation, and networks). Including studies published in languages other than English can arguably minimize bias but may be harder to retrieve (Toews et al., 2017). If included, it is important for the reviewers to consider, as the process is interpretive, how to preserve conceptual meanings and map the themes when translating from different languages. Equally, choosing a specified time frame or an “all-years” approach depends on what is pragmatic and meaningful for your review question and the chosen databases (e.g., on Web of Science the earliest index period is 1950). Whatever the approach, it would be necessary to justify the decision on which the literature search was based. For example, in one of our reviews (Soilemezi et al., 2017), the date of a new social care legislation was used as the start date for the searches and we included German and Greek studies (in addition to English) as members of the review team were fluent in these languages.

Saving your results. Having identified all the relevant studies, the final list is typically imported either into reference software (e.g., EndNote, Mendeley, RefWorks, Zotero) designed to manage large quantities of references or to specialist systematic review software such as EPPI-Reviewer, DistillerSR, Covidence, and Qualitative Assessment Review Instrument (QARI; Pearson, 2004). The use of such software is helpful to keep track and file the imported references accordingly. A main challenge identified by Saini and Shlonsky (2012) in this phase, that also echoes our experiences, is the lack of databases’ flexibility to transfer the citations to reference management software. The smoothest transfers occur when the database and software are from the same provider (e.g., EndNote with Web of Science, Mendeley and RefWorks with Science Direct). Some databases have a limit on how many citations that can be imported per time (e.g., 50 hits on Web of Science, 100 on EBSCO), potentially resulting in a frustrating and time-consuming process, especially where the search results contain over 1,000 hits. In some databases (e.g., British Architecture Library Catalogue), only manual import is possible, while in

Social Care Online it is only possible to import the top 500 results. EndNote searches for full papers (if available) and automatically saves them without the need to manually search for them later. It is advisable to be aware of the idiosyncrasies of each database and software before deciding which ones to include and to save the searches on each database to be able to rerun and update, if needed. This is not to say that you should avoid particular databases that may be very important for your topic; it is to warn new reviewers that additional action(s) may be needed if you use an “awkward” database (e.g., paste the records into a word document and then upload them to your reference manager).

Once all of your references are imported into the software of your choice, the next step is retrieval and removal of duplicates. Sometimes, automated removal of duplicates is only partially successful and further manual removal may be necessary (Rathbone, Carter, Hoffmann, & Glasziou, 2015). One option can be to copy the records to a second software program and attempt further detections of duplicate records there. However, this might also cause problems, as the software might not be able to retain the same unique reference ID numbers, and therefore accurately track references. In any case, it is essential to record the number of duplicates (before and after the removal of duplicates) and be aware that databases may have different ways of recording citations (e.g., variations in page numbers, author details), and hence they may not always be successful in retrieving the duplicate records.

Screening of Identified Studies

The second step in a systematic review is to screen the studies identified by the searches to ensure they can potentially answer the review question. This step involves two challenges: the number of reference screening stages, and the decision of whether screening of each reference should be performed by one or more reviewers.

Screening stages. This exercise typically starts by screening titles and abstracts (where available) simultaneously. Alternatively, screening the titles of eligible references, eliminating irrelevant ones, and later screening the abstracts of those that are thought relevant is also possible. Following exclusion of the abstracts that do not meet the inclusion criteria, reviewers move on to screen the full papers. This three-stage process follows the quantitative review process. Research by Mateen, Oh, Tergas, Bhayani, and Kamdar (2013) indicated that although screening simultaneously by titles and abstract is a more accurate strategy, the screening title-only approach may be more efficient in reducing the time required to get to the final included papers. Although this may be a quicker screening approach for quantitative reviews, our experience with screening qualitative studies showed that this was not effective. This is because often titles are not clearly identified as qualitative papers, not presenting sufficient information regarding the research aims, and thus potentially relevant papers may be missed and not taken through to the abstract stage (Flemming

& Briggs, 2007; Flemming & McInnes, 2012; Jones, 2004; Kavanagh et al., 2012). A two-stage process (screening title/abstract together and moving to screening with full texts) could be a more pragmatic and rigorous strategy which, if done systematically, may potentially save the reviewers time from adding later missed studies. Reviewers should consider whether the article addresses the phenomenon and research question in mind and was published in the time period agreed in the protocol, language eligibility, population of interest, and if the study design is identified (Porritt, Gomersall, & Lockwood, 2014).

If reviewers choose to follow the title-only screening approach, it is advisable to remain inclusive and apply more strict criteria once they get to read the abstracts. Flemming and Briggs (2007) found that many papers were identified by the reviewers' own interest and knowledge of the literature even though these papers were mistakenly excluded in the title phase, which was also the case from our experience. Nonetheless, due to inadequate information about the study designs in the qualitative titles and abstracts, the decision whether to include or exclude is often only possible after the retrieval of full texts (Jones, 2004). In our experience, only after assessing the full article, it was possible to ascertain if some papers contained relevant data for the review, which at first did not appear relevant. Similarly, it has been suggested that the first-level screening criteria is that the articles must reflect the research question and objective(s), not the methodology used (Kavanagh et al., 2012; Saini & Shlonsky, 2012).

Double or single screening. Double screening is regarded the best practice for systematic reviews to minimize bias and chances of missing relevant papers. However, when the search results are relatively low in number (e.g., only a few hundred), and/or the time scheduled for completion is very limited, the lead reviewer may screen all papers and bring the final selection to the coreviewers for evaluation and extraction. This strategy, however, has some limitations, namely, the possibility to exclude relevant papers and to introduce bias. In addition, it has been argued that the value of two reviewers in qualitative synthesis is not to reach consensus or verify data but to identify multiple perspectives, that is, for dissonance (Booth et al., 2016). From our experience, when more than two reviewers are involved in the screening phase, the results are more rigorous (with disagreements becoming part of the sensitization process), the team remains engaged with the topic and procedure, there is more support for the main reviewer to be critical, transparent about decisions, and enthusiastic for what might be a long task! Double screening may prolong progression but the advantage would be that the inclusion/exclusion criteria would be tighter and clear to all reviewers. Alternative ways to screen could be (a) the main reviewer to screen all eligible references and for the remaining team to screen equally divided portions of eligible references, or a percentage of them, and (b) for one reviewer to screen all references and another to screen only the excluded ones.

In our experience, it is better to pilot the screening criteria and process to ensure that all reviewers (particularly new reviewers) are able to apply them consistently. It is likely that reviewers may face uncertainty over the eligibility of studies and this may result in many "unclear" papers. The decision then to include/exclude will depend on the protocol, the full text, or rules the reviewers may set up as they go along (as long as a track of all decisions is recorded and adhered to by all reviewers).

Eligibility of Studies and Methodologies

Following the screening titles and abstracts for inclusion, the next step is to assess the studies for their eligibility and quality.

Assessing eligibility of full-text papers. Whether screening the full papers electronically or using a hard copy, the challenge here is to look out for linked studies and to get an electronic or printed version of all potentially relevant papers. For nonopen access references, for papers not in journals subscribed to the organizations' library, or papers that predate electronic versions, requesting interlibrary loans (ILL) is the common option. The challenges here could be (a) some ILL services are expensive to use and (b) only one reviewer has the copyright to access this. An alternative solution would be to contact the lead author(s) to request a copy, if permitted. Nowadays, social media platforms (e.g., ResearchGate) can make it easier to find researchers (if they have a profile), although a (speedy) response is not guaranteed. It is also helpful to think strategically about how to deal with missing data (e.g., set realistic deadlines for hearing back from authors before deciding whether to exclude a paper) depending on time and resources available.

When assessing full-text papers, two reviewers always complete the full-text screening, even if previous stages (title and abstract) were not double screened. A study only has to fail one criterion to be excluded. At this stage, it is expected that the reason(s) for exclusion will be documented, and reported on the PRISMA flowchart. If a study failed many criteria, the primary reason for exclusion is noted. However, in a more inclusive approach, it might be decided to include studies that, although their primary focus would not directly answer the review question, they might have useful data that can contribute to the overall knowledge base (Hannes & Pearson, 2012)

Appraising quality. Conducting a transparent appraisal requires identifying and exploring whether the eligible studies are fit for purpose before proceeding to final synthesis of the data. Critical appraisal or quality assessment is important as studies can be poorly conducted or reported and findings may be unreliable, which may bias the review outcomes. This step can assure credibility, rigor, and trustworthiness of the synthesis as well as aid transparency of the decisions made (Paterson, 2012; Porritt et al., 2014; Spencer, Ritchie, Lewis, & Dillon, 2003). With a multiplicity of qualitative approaches and a striking proliferation of over 100 structured quality assessment checklists (Dixon-Woods et al., 2006; Noyes et al., 2013; Saini &

Shlonsky, 2012), quality appraisal has been a topic of debate and critique with some reviewers suggesting that it should not be done and others arguing that it is an important filtering step that adds value to the review if account for the diversity of qualitative methods (Atkins et al., 2008; Dixon-Woods et al., 2007; Lewis et al., 2015; Pope & Mays, 2006b).

Searching for the best tool to use can be confusing, but should encompass the key markers of the quality of qualitative research. If necessary, an existing tool may be expanded by adding questions and indicators that are relevant to the appraisal (Spencer et al., 2003). Common appraisal tools are the QARI (<https://joannabriggs.org/assets/docs/sumari/ReviewersManual-2014.pdf>), the Transparency Accuracy Purposively Utility Proprietary Accessibility (TAPUPA; <https://www.scie.org.uk/publications/knowledgereviews/kr03.asp>) framework mainly used in social care, the Quality Framework developed by the Cabinet Office (https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/498321/Quality-in-qualitative-evaluation_tcm6-38739.pdf), the Critical Appraisal Skills Programme (CASP; <http://www.casp-uk.net/casp-tools-checklists>) checklist, and the Evaluation Tool for Qualitative Studies (ETQS; http://usir.salford.ac.uk/12970/1/Evaluation_Tool_for_Qualitative_Studies.pdf) as well as ones individually tailored by different reviewers to fit their needs.

To our knowledge, there is no formal guidance on how to choose an appraisal instrument. Although what constitutes a concept of quality remains debatable (Noyes et al., 2013), and despite studies reporting checklists and criteria for conducting good qualitative research (Cohen & Crabtree, 2008; Garside, 2014; Tong, Sainsbury, & Craig, 2007), choosing the right appraisal tool(s) can be a bewildering task for a new reviewer. Using a working example, Dixon-Woods et al. (2006) question the usefulness of the critical appraisal templates, as often reviewers cannot reach consensus on the quality of studies, reported findings, and the relevance of the topic. In another study, checklists or structured approaches, used to appraise studies for inclusion in a systematic review, did not produce higher agreement between reviewers when compared with unprompted judgment (Dixon-Woods et al., 2007). In a comparative study by Hannes, Lockwood, and Pearson (2010), it was found that the ETQS provides more detailed instructions on how to apply the evaluation criteria, the JBI tool (although does not address external validity or relevance) is the most coherent, whereas the CASP tool may be less sensitive to validity but is a popular tool to use for novice researchers.

Carroll and Booth (2015) argue that a combination of instruments should be considered, and the choice of which one(s) to use should be based on the review team's expertise (e.g., experience in primary qualitative research, in theoretical/philosophical perspectives) alongside the requirements of review context and question. As a standard form may not be fit to all synthesis approaches, the review team needs to reflect on these factors before deciding which is the best tool to use. In our experience, despite the use of the several quality appraisal tools, several difficulties may be present regarding the decision

to include or exclude studies and how many reviewers to involve. For example, the QARI instrument (Pearson, 2004), requires the evaluation of the philosophical perspectives of the qualitative studies and their congruence but does not offer sufficient guidance on this. Also, without the papers being assessed by *at least two* reviewers independently, the review cannot progress. Typically, quality assessment of papers is double-checked by two independent reviewers for accuracy and mutual agreement, although this depends largely on time and resources available.

Regardless of the tool(s) used, it is worth considering two parameters: (a) the cutoff point for inclusion or thresholds of quality/bias (such as high, medium, and low) and (b) where the paper was published. Some appraisal tools ask the reviewers to score each study against some criteria, which means that if these criteria are not all reported, the studies will score low. However, the use of numerical quality scores in systematic reviews has been criticized and it has become increasingly common not to use scores or a strict cutoff criterion because this is associated with judging the quality of the written report rather than the uniqueness of the research process itself (Atkins et al., 2008; Sandelowski & Barroso, 2002). Not scoring high on all aspects does not necessarily mean that the study was of poor quality, it might be that some aspects were merely not reported (Atkins et al., 2008). Authors often have to adhere to strict word limits, to peer review, or editors' suggestions and therefore some information might be missing from the report (not necessarily from the study itself) to enable a full appraisal (Sandelowski & Barroso, 2002). Rather than excluding studies at the outset, it has been argued that the reviewer should use the tools as part of exploration and judge each paper's contribution to the synthesis based upon the relevance, the objectives, the theoretical sensitivity in relation to the review aims, and credibility (Atkins et al., 2008; Dixon-Woods et al., 2006, 2007; Saini & Shlonsky, 2012; Spencer et al., 2003). This judgment call often depends on the reviewers' disciplines and/or their methodological preferences (Jones, 2004).

Additionally, it is argued that it can often be difficult to judge the quality of each research study for the inclusion in the synthesis due to their significant differences in theoretical perspectives, methodologies, and diverse epistemological assumptions (Erwin et al., 2011). For example, reviewers using CIS typically do not conduct a formal quality evaluation as they are focused on the papers' relevance to the review question and not the methodological aspects (Flemming & McInnes, 2012). Regardless of the review approach (including only studies that are in line with the epistemological underpinning of the method of synthesis or open to inclusion of diversity of study designs), tools that require the assessment of philosophical perspectives can prove to be complicated for new reviewers. This requires a level of expertise in the team and the need for training for new reviewers in order to assess for congruence between the studies (Booth et al., 2016). For example, the CASP tool has questions focused on the aims of the research but not the philosophical approach, whereas the QARI tool has questions on the congruence between philosophical perspectives and methodologies of

the included primary qualitative studies. Our experience using QARI is that this could be a challenging tool to use for a junior reviewer with limited experience in different philosophical perspectives, which could lead to disagreements and delays. Although, the disciplinary background of the review team can be beneficial, it is essential to consider team's expertise for each task and importantly to be clear about the ultimate aim of the synthesis (interpretive, aggregative, or integrative).

To exclude or not to exclude? Despite conducting quality appraisals, many reviewers decide not to exclude "inferior" studies (Noyes et al., 2008). Reviewers may include these studies if they report useful and authentic accounts of a phenomenon despite being poorly reported. The quality of reporting qualitative studies has improved in recent years, but this may be significantly poorer in older studies. Studies published in qualitative journals, perhaps due to their more generous word limits, report more information on the research process, rather than studies published in medical journals (Atkins et al., 2008; Jones, 2004). Thus, it is important to bear in mind the field of the synthesis when judging the quality of studies and consider other forms of clarification, for example, contact the authors. On the other hand, other researchers argue not to include studies of poor quality as it would bias their findings and limit recommendations (Pearson, 2004).

For reviewers who decide to be inclusive, it is suggested to run a sensitivity analysis by removing the "low-quality" studies, as can be done in quantitative reviews (Carroll & Booth, 2015; Dixon-Woods et al., 2007; Noyes et al., 2008). The rationale for this is that it is anticipated that poor-quality studies would contribute minimally to the formation of synthesized themes and final recommendations (Britten & Pope, 2012; Carroll & Booth, 2015). Reflecting on how critical appraisal findings should be used in qualitative reviews, Carroll and Booth (2015) commented that "*there is evidence that the exclusion of studies can adversely affect the generalisability of a review and synthesis*" (p. 152). It is therefore recommended that reviewers run a sensitivity analysis as a usual practice, and/or to assess the quality of included papers later on in the process when they can clearly judge the value of each paper to the synthesis and only then decide which ones to exclude based on "*the conceptual richness rather than methodological quality*" (Britten & Pope, 2012, p. 56).

Inclusion, Synthesis, and Reporting

At the final stage, reviewers need to extract data from the eligible studies, decide how to synthesize the findings, and write a report.

Extracting data. Typically, the reviewers need to extract necessary information from each included paper, namely, descriptive data (regarding the participants, information on the study type and characteristics, location, setting, year, main topic, etc.), methods, type of analysis, findings, and original quotations

(Munn, Tufanaru, & Aromataris, 2014). The level of detail depends on the review's aims. The extraction can be performed using standardized data extraction forms (e.g., JBI-QARI) or the team may decide to develop (or adapt an existing form) for the purpose of the given review, which can be saved on a Word, Excel, or Google Forms. This process is straightforward once the team agrees on, and records instructions of, what data to extract (to form the data set for the synthesis) and how to deal with missing information. This is specifically important when two reviewers extract data independently in order to keep consistency and compare findings. The number of reviewers to be involved in this stage depends on the research approach, the tool used (e.g., JBI-QARI requires two reviewers), time, and other resources. Double data extraction is considered the golden standard, but it is not always the case. Often one reviewer is tasked with the whole data extraction and other(s) contribute to a percentage (e.g. 10–20% of random sample of the included papers) to provide some quality reassurance regarding the adequate data extraction.

It is common to initiate the data extraction in tandem with the quality appraisal or it can be done as a separate stage. As in previous reviews (Britten & Pope, 2012), our experience also indicates that it is more practical to do these two activities together as it is a more convenient and time efficient process. However, this is not always possible (e.g., when using QARI software as these tasks are separate).

Synthesizing the findings. Choosing the right analytical approach guides the synthesis of qualitative research and usually depends on the research question and scope of the review, available evidence (number and type of data: homogeneous or heterogeneous), team size, expertise and commitment, and other resources (Booth et al., 2016; Britten & Pope, 2012; Noyes et al., 2008; Ring et al., 2011). Some approaches allow the inclusion of different qualitative methods, but some argue that they should not be combined. It is beyond the scope of this article to present and compare different approaches as this has been reported previously (Barnett-Page & Thomas, 2009; Booth et al., 2016; Dixon-Woods, Agarwal, Jones, Young, & Sutton, 2005; Hannes & Lockwood, 2012; Ring et al., 2011). As there are no rules on what is considered an adequate (minimum or maximum) number of studies (Lewin et al., 2015), the challenge in this stage is for reviewers to consider whether they can produce a meaningful synthesis with the included studies. Booth et al. (2016) reported that around a dozen papers could offer an optimal trade-off between richness of data and feasibility of the review, although there have been reviews that included just three studies. Some aggregative approaches may handle large number of papers although some interpretive approaches may benefit from a small number of studies (Booth, 2016). Undoubtedly, including a large number of studies may be unmanageable and reviewers may choose to refine the review question (e.g., population, condition). On the other hand, including a small number of studies may not produce a meaningful result and it may indicate that the team need to expand the review question (Lewin et al., 2015). However,

Lewin and colleagues (2015) noted that fewer but more conceptually rich studies might contribute more than a large number of thin studies. Thus, reviewers need to consider not only the number but also the richness of the studies.

Regardless of the chosen approach, each with unique challenges, there will be a process of collating evidence from individual studies to form new findings on the same topic. Methods of analysis usually have (to a smaller or larger degree) some level of immersion, categorization, combining and making sense of the data, and developing new themes to be able to reach new understandings, conclusions, and/or recommendations. It is also common that the findings of the primary included studies will become your coding data. What constitutes findings will depend on the approach and reviewers to define and explain. For example, in one of our reviews (Soilemezi et al., 2017) using thematic synthesis, findings comprised all of the text included in the findings section, including quotations. Using the JBI approach in another of our reviews (Linceviciute, Dewey, & Kilburn, 2013), we did not typically include non-supported (without direct quotations from the participants in the primary studies) statements, as only participants' words (i.e., quotations) are considered findings and used as the coding data.

It is often a challenge to decide the selection and length of the text and/or quotation to extract and analyze (Atkins et al., 2008; Gallacher et al., 2013). What is crucial at this stage is to create a rigorous opportunity to extract insights that might not be possible on the basis of single studies alone. Generally, the first step is to create initial codes according to a given analytical approach. As the analysis progresses, it might be that some findings form a code, or they might become a code combined with other findings. It might also require for some codes to be collapsed in order to form another code or (sub)theme. It can be challenging to extract and combine themes from several studies to come to a new finding and/or conclusion (Ring et al., 2011). What is crucial though is to document clearly the steps followed. It is often reported that by synthesizing their data, researchers may adapt their method to combine other approaches, without necessarily explaining how these amalgamations resulted and why (Paterson, 2012). To overcome these challenges, it is advisable to remain focused on the review question and close to the original articles for context and clarifications, especially if the included studies vary considerably (Atkins et al., 2008).

When considering the relevance of the data from different studies, the reviewers may need to think about two factors: geographical location and time periods. If data are synthesized representing findings from different countries and continents, the reviewers may need to question how these may be relevant and applicable to what population and how this may inform their conclusions and recommendations. Equally important would be to think how to analyze studies from different time frames and how applicable the findings would be to inform current practice. One option would be to apply a strict time division of papers using a meaningful date (e.g., year of a new legislation, new intervention) and conduct two analyses: one

including the papers before and one after that date. The other option would be to include all papers to form the main themes but remain mindful of potential differences due to the time frame. In this case, the reviewers will need to make a statement and discuss about how the themes from different time frames are relevant (or not any more) to answer your question and guide clinical practice. This decision will depend on the team's expertise, the research question, and the number of eligible papers.

Review software (e.g. QARI, EPPI-Reviewer) can be of great use for qualitative reviewers alongside typical qualitative software (e.g. NVivo 12, Atlas-ti 8, MAXQDA 2018). The merits of using review software are that they can make the analytical process more manageable, there is a trail of decisions and transparency in the analytical process and they allow coding, organizing themes, easily retrieving sections when later writing the results. Both EPPI-Reviewer and QARI are web-based and they have functions and tools to support an audit trail of all stages of reviewing (saving records, screening, quality assessment tool, data extraction, data synthesis, and reporting), whereas typical qualitative software can only be used in the final stages of analysis and reporting (Hannes & Pearson, 2012). EPPI-Reviewer requires a subscription fee for each user and thus restricted funding may be a barrier in using it (especially for long-term reviews and with many reviewers involved as this will increase the costs). QARI software comes with an annual fee to be paid separately by all reviewers but also allows institutional use, in which case individual reviewers do not have to pay separately. One of QARI software's limitations, however, is that it does not allow reviewers to construct sub-categories and only limits the findings into main categories (Hannes & Pearson, 2012).

Writing up the findings. The final task is writing up the process and presenting the findings of the review in a paper, chapter, or report. Reporting the process of the review can be straightforward if the reviewers have kept systematic and transparent documentation of all milestones and decisions made throughout the review journey to demonstrate rigor, credibility, and reflection over the process and methodology (Erwin et al., 2011). The most important are documenting the search strategy, the final search terms used in each database, the final number of citations hits in each database, the number of the excluded studies, the list of the included studies, the data analysis strategy, and the method of analysis. Any deviations from the protocol will need to be explained and justified.

As reviews are often used to building a bridge from research to practice, it is expected that the reviewers write in an accessible, unbiased, and usable format to inform different audiences, provide enough information for the readers to understand and decide whether and how to apply the findings (Chalmers & Glasziou, 2009; Erwin et al., 2011). The plain and transparent reporting that provides important information on selection and publication bias adds to the credibility of the findings (Robertson-Malt, 2014). Depending on the findings and approach, it might be appropriate to present the findings

in mind maps, tables, charts, figures, or plain text. Reviewers are advised to follow well-established guidelines such as the ENTREQ (Tong et al., 2012), the EQUATOR network, the PRISMA flowchart, and GRADE-CERQual approach. In the writing stage in particular, the use of computer software can be invaluable to trace the primary findings back to the included studies and the decisions made in theme formulation.

Perhaps the biggest challenge at this stage is to draw the conclusions and recommendations for practice and research based on the synthesized findings and to remain specific about the claims that can be made. In reviews when qualitative and quantitative data are integrated at the final stage, the findings from both components can be presented either in a matrix, tabular, narrative, or graphical form or in a conceptual framework with an independent reviewer joining the two components (Booth et al., 2016). Also, it is important to identify and report the negative cases (if any), limitations, and barriers in the synthesis processes: identification, screening, eligibility, and analysis (Lewin et al., 2015; Robertson-Malt, 2014). This will help the reader (researcher, practitioner, and policy maker) to draw conclusions with more confidence. In some cases, it might be appropriate to involve an advisory group of public members (e.g., service users, clinicians) to validate the interpretations and relevance of the review. However, this may prove difficult if public members are not familiar with constructed themes and unable to assess the relevance (Lewin et al., 2015). Saini and Shlonsky (2012) argued the importance of reporting whether the review findings could have applicability and transferability beyond the population studied. Unlike quantitative reviews that typically are updated every few years (depending on the field and progress of evidence), this is not usually required in qualitative reviews.

Discussion

Systematic reviewing of research is inevitably demanding and time-consuming with every review having its own challenges and every researcher having a different set of skills and resources to utilize and deal with these. In this article, we tried to present a brief summary of the process and highlight some of the potential challenges that new reviewers may face in the course of this process. The aim was to present examples and pragmatic options rather than portray exact actions in order to demystify difficulties and support reviewers to complete the review within the time and resources available. Exploring these uncertainties can enable reviewers to address difficulties effectively, optimize their choices, and reflect on their practices.

Given the diversity in methods and approaches, conducting a systematic review requires flexibility, clinical and/or academic knowledge, and being able to justify the decisions and disagreements along the way. As suggested by other reviewers (Gallacher et al., 2013; Ring et al., 2011), we also support the idea that two reviewers should be involved in all stages, one with previous reviewing experience (even in quantitative reviews) and one with qualitative research expertise, and, if possible, one with expertise in the topic being

reviewed. This synergetic reviewing and synthesizing will bring different perspectives, assist with transparency, minimize bias, and add validity and richness to the findings. However, we acknowledge that qualitative research is essentially subjective and it is unlikely that even the most experienced reviewers will always reach a consensus when screening papers despite clear protocols and checklists or produce exactly the same themes (Gallacher et al., 2013; Pearson, 2004).

Despite the ongoing developments in the automation of systematic reviews (e.g., the Cochrane Crowd and the development of “Screen for Me” service, machine learning to screen references for inclusion/exclusion), some tasks remain largely manual (Tsafnat et al., 2014). Until technology and databases are developed for synthesizing (e.g., allow hundreds of hits to be extracted, software to include sensitivity analysis) and develop further existing or new software to support the process, the work of the reviewers is likely to be long and demanding. However, improving some aspects further can make qualitative synthesis a more rewarding process. For example, in the identification phase, researchers and librarians should receive more training in qualitative searches for systematic reviews. The current advice is to remain overinclusive (where appropriate) to eliminate risk of missing out potential relevant records (Shaw et al., 2004). Reliance on medical databases (e.g., MEDLINE) is not enough; other databases should also be searched (social, nursing, psychological, and educational) and especially CINAHL, which has good indexes for qualitative methodology (Evans, 2002). Ongoing development of methodological filters for different databases is also needed to improve search strategies (Booth, 2016). For the facilitation of results’ retrieval, it is convenient to use software to manage and record the citations. Clear documentation of the search strategy, number of hits, and duplications is vital to ensure robustness and reproducibility. When it comes to the screening phase, authors of qualitative papers should produce well-structured titles and abstracts (Atkins et al., 2008; Jones, 2004; Shaw et al., 2004) that are appropriately identified on title and indexed under qualitative terms in order to be easily retrieved and included in future reviews. In the eligibility phase, qualitative authors should state clearly all criteria that an appraisal tool would require to enable a rigorous quality assessment. Each review has unique quality issues and decisions, and there are cases where the reviewers might decide to apply strict criteria, or be inclusive, or even decide not to carry on a systematic quality appraisal at all, perhaps for reasons of limited records or authoring some of the included papers (Hannes & Pearson, 2012).

In the synthesis phase, clear analytical steps based on a preferred approach should be followed and reported. The reviewers have the challenging work to extract insights from single studies and critically interrogate them, without “removing” the original work that made the included studies diverse, nuanced, and meaningful. More guidance on choosing and reporting the right methods now exist to assist new reviewers (see RAMESES Projects [http://www.ramesesproject.org/Home_Page.php] and Booth et al., 2016). We take the

view that sometimes it is best to decide the analysis once inclusion screening, data extraction, and critical appraisal are finalized in order to decide whether and what type of synthesis is possible (and meaningful). Although this approach is not suitable for all review questions, it is typically used in quantitative reviews where, if statistical analysis is not possible, the usual practice is to produce a narrative summary (Pearson, 2004). Perhaps a more flexible approach for qualitative reviewers could be possible to enable the review team to produce the best answer within time and resources available. Finally, synthesizing the results could be strengthened by validation with the population under question to ensure the findings are relevant and applicable to practice.

In conclusion, qualitative synthesis is a thought-provoking and rewarding process and if planned carefully, it can be less stressful, unpredictable, and resource intensive for reviewers. Careful planning involves being systematic in not only the methods but also planning the management of the process, which is often underestimated. Dealing with technique as well as substance is important to generate new knowledge and offer greater understanding in the field.

Acknowledgments

We are grateful to the reviewers for their helpful comments. We would also like to thank Prof. Alan Costall, Dr. Muhammad Hossain, Dr. Jonathan Shepherd and Louise Baxter for their ideas and useful comments.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Dia Soilemezi  <http://orcid.org/0000-0001-8831-9419>

References

- Atkins, S., Lewin, S., Smith, H., Engel, M., Fretheim, A., & Volmink, J. (2008). Conducting a meta-ethnography of qualitative literature: Lessons learnt. *BMC Medical Research Methodology*, *8*, 21. Retrieved from <http://doi.org/10.1186/1471-2288-8-21>
- Barnett-Page, E., & Thomas, J. (2009). Methods for the synthesis of qualitative research: A critical review. *BMC Medical Research Methodology*, *9*, 59. Retrieved from <http://doi.org/10.1186/1471-2288-9-59>
- Booth, A. (2016). Searching for qualitative research for inclusion in systematic reviews: A structured *methodological review*. *Systematic Reviews*, *5*, 1–23. doi:10.1186/s13643-016-0249-x
- Booth, A., Noyes, J., Flemming, K., Gerhardus, A., Wahister, P., van der Wilt, G. J., . . . Rehfuss, E. (2016). *Guidance on choosing qualitative evidence synthesis methods for use in health technology assessments of complex interventions [online]*. Retrieved from <http://www.integrate-hta.eu/downloads/>
- Britten, N., & Pope, C. (2012). Medicine taking for asthma: A worked example of meta-ethnography. In K. Hannes & C. Lockwood (Eds.), *Synthesizing qualitative research* (pp. 41–57). Chichester, England: John Wiley & Sons Ltd.
- Butler, A., Hall, H., & Copnell, B. (2016). A guide to writing a qualitative systematic review protocol to enhance evidence-based practice in nursing and health care. *Worldviews on Evidence-Based Nursing*, *13*, 241–249.
- Carroll, C., & Booth, A. (2015). Quality assessment of qualitative evidence for systematic review and synthesis: Is it meaningful, and if so, how should it be performed? *Research Synthesis Methods*, *6*, 149–154. Retrieved from <http://doi.org/10.1002/jrsm.1128>
- Chalmers, I., & Glasziou, P. (2009). Avoidable waste in the production and reporting of research evidence. *Lancet*, *374*, 86–89. Retrieved from [http://doi.org/10.1016/S0140-6736\(09\)60329-9](http://doi.org/10.1016/S0140-6736(09)60329-9)
- Cohen, D. J., & Crabtree, B. F. (2008). Evaluative criteria for qualitative research in health care: Controversies and recommendations. *Annals of Family Medicine*, *6*, 331–339. Retrieved from <http://doi.org/10.1370/afm.818>
- Cooke, A., Smith, D., & Booth, A. (2012). Beyond PICO: The SPIDER tool for qualitative evidence synthesis. *Qualitative Health Research*, *22*, 1435–1443. doi:10.1177/1049732312452938
- Dixon-Woods, M., Agarwal, S., Jones, D., Young, B., & Sutton, A. (2005). Synthesising qualitative and quantitative evidence: A review of possible methods. *Journal of Health Services Research & Policy*, *10*, 45–53. Retrieved from <http://jhsrp.rsmjournals.com/cgi/content/abstract/10/1/45>
- Dixon-Woods, M., Agarwal, S., Young, B., Jones, D., & Sutton, A. (2004). *Integrative approaches to qualitative and quantitative evidence*. London, England: NHS Health Development Agency.
- Dixon-Woods, M., Bonas, S., Booth, A., Jones, D. R., Miller, T., Sutton, A. J., . . . Young, B. (2006). How can systematic reviews incorporate qualitative research? A critical perspective. *Qualitative Research*, *6*, 27–44. Retrieved from <http://doi.org/10.1177/1468794106058867>
- Dixon-Woods, M., Sutton, A., Shaw, R., Miller, T., Smith, J., Young, B., . . . Jones, D. (2007). Appraising qualitative research for inclusion in systematic reviews: A quantitative and qualitative comparison of three methods. *Journal of Services Research & Policy*, *12*, 42–47. Retrieved from <http://hsr.sagepub.com/content/12/1/42.short>
- Erwin, E. J., Brotherson, M. J., & Summers, J. A. (2011). Understanding qualitative metasynthesis. *Journal of Early Intervention*, *33*, 186–200. Retrieved from <http://doi.org/10.1177/1053815111425493>
- Evans, D. (2002). Database searches for qualitative research. *Journal of the Medical Library Association*, *90*, 290–293. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC116400/>
- Flemming, K., & Briggs, M. (2007). Electronic searching to locate qualitative research: Evaluation of three strategies. *Journal of Advanced Nursing*, *57*, 95–100. Retrieved from <http://doi.org/10.1111/j.1365-2648.2006.04083.x>
- Flemming, K., & McInnes, E. (2012). The use of morphine to treat cancer related pain: A worked example of critical interpretive synthesis. In K. Hannes & C. Lockwood (Eds.), *Synthesizing*

- qualitative research* (pp. 59–82). Chichester, England: John Wiley & Sons Ltd.
- Gallacher, K., Jani, B., Morrison, D., Macdonald, S., Blane, D., Erwin, P., . . . Mair, F. S. (2013). Qualitative systematic reviews of treatment burden in stroke, heart failure and diabetes: Methodological challenges and solutions. *BMC Medical Research Methodology*, *13*, 10. Retrieved from <http://doi.org/10.1186/1471-2288-13-10>
- Garside, R. (2014). Should we appraise the quality of qualitative research reports for systematic reviews, and if so, how? *Innovation*, *27*, 67–79. Retrieved from <https://doi.org/10.1080/13511610.2013.777270>
- Hannes, K. (2011). Critical appraisal of qualitative research. In J. Noyes, A. Booth, K. Hannes, J. Harris, S. Lewis, & C. Lockwood (Eds.), *Supplementary guidance for inclusion of qualitative research in Cochrane systematic reviews of interventions*. Retrieved from <http://cqrmg.cochrane.org/supplemental-handbook-guidance>
- Hannes, K., & Lockwood, C. (2012). *Synthesizing qualitative research: Choosing the right approach*. Chichester, England: John Wiley & Sons Ltd.
- Hannes, K., Lockwood, C., & Pearson, A. (2010). A comparative analysis of three online appraisal instruments' ability to assess validity in qualitative research. *Qualitative Health Research*, *20*, 1736–1743. Retrieved from <http://doi.org/10.1177/1049732310378656>
- Hannes, K., & Pearson, A. (2012). Obstacles to the implementation of evidence-based practice in Belgium: A worked example of meta-aggregation. In K. Hannes & C. Lockwood (Eds.), *Synthesizing qualitative research: Choosing the right approach* (pp. 21–39). Chichester, England: John Wiley & Sons Ltd.
- Jenkins, M. (2004). Evaluation of methodological search filters—A review. *Health Information and Libraries Journal*, *21*, 148–163. Retrieved from <http://doi.org/10.1111/j.1471-1842.2004.00511.x>
- Jones, M. L. (2004). Application of systematic review methods to qualitative research: Practical issues. *Journal of Advanced Nursing*, *48*, 271–278. Retrieved from <http://doi.org/10.1111/j.1365-2648.2004.03196.x>
- Kavanagh, J., Campbell, F., Harden, A., & Thomas, J. (2012). Mixed methods synthesis: A worked example. In K. Hannes & C. Lockwood (Eds.), *Synthesizing qualitative research: Choosing the right approach* (pp. 113–136). Chichester, England: John Wiley & Sons Ltd.
- Lewin, S., Glenton, C., Munthe-Kaas, H. M., Carlsen, B., Colvin, C. J., Gulmezoglu, M., . . . Rashidian, A. (2015). Using qualitative evidence in decision making for health and social interventions: An approach to assess confidence in findings from qualitative evidence syntheses (GRADE-CERQual). *PLoS Medicine*, *12*, 1–18. Retrieved from <https://doi.org/10.1371/journal.pmed.1001895>
- Linceviciute, S., Dewey, A., & Kilburn, S. (2013). Exploring the perceptions and experiences of older adults with asthma: A synthesis of qualitative asthma studies in older population. PROSPERO 2013 CRD42013006849. Retrieved from http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42013006849
- Mateen, F. J., Oh, J., Tergas, A. I., Bhayani, N. H., & Kamdar, B. B. (2013). Titles versus titles and abstracts for initial screening of articles for systematic reviews. *Clinical Epidemiology*, *5*, 89–95. Retrieved from <http://doi.org/10.2147/CLEP.S43118>
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., . . . Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, *4*, 1. Retrieved from <http://doi.org/10.1186/2046-4053-4-1>
- Munn, Z., Tufanaru, C., & Aromataris, E. (2014). JBI's systematic reviews: Data extraction and synthesis. *American Journal of Nursing*, *114*, 49–54. doi:10.1097/01.NAJ.0000451683.66447.89
- Noyes, J., Hannes, K., Booth, A., Harris, J., Harden, A., Popay, J., . . . Pantoja, T. (2013). Chapter QQ: Qualitative and implementation evidence and Cochrane reviews. In J. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions* (pp. QQ1–26). The Cochrane Collaboration. Retrieved from <http://qim.cochrane.org/supplemental-handbook-guidance>
- Noyes, J., Popay, J., Pearson, A., Hannes, K., & Booth, A. (2008). Qualitative research and Cochrane reviews. In J. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions* (pp. 571–591). Chichester, England: John Wiley & Sons.
- Paterson, B. L. (2012). “It looks great but how do I know if it fits?” An introduction to meta-synthesis research. In K. Hannes & C. Lockwood, *Synthesizing qualitative research* (1st ed., pp. 1–20). Chichester, England: John Wiley & Sons Ltd.
- Pearson, A. (2004). Balancing the evidence: Incorporating the synthesis of qualitative data into systematic reviews. *JBI Reports*, *2*, 45–64. Retrieved from <http://doi.org/10.1111/j.1479-6988.2004.00008.x>
- Pope, C., & Mays, N. (2006a). *Qualitative research in health care* (3rd ed.). Oxford, England: Blackwell.
- Pope, C., & Mays, N. (2006b). Synthesizing qualitative research. In C. Pope & N. Mays (Eds.), *Quality research in health care* (3rd ed., pp. 142–152). Oxford, England: Blackwell.
- Porritt, K., Gomersall, J., & Lockwood, C. (2014). Study selection and critical appraisal. *The American Journal of Nursing*, *114*, 47–52. doi:10.1097/01.NAJ.0000450430.97383.64
- Rathbone, J., Carter, M., Hoffmann, T., & Glasziou, P. (2015). Better duplicate detection for systematic reviewers: Evaluation of systematic review assistant-deduplication module. *Systematic Reviews*, *4*, 6. Retrieved from <http://doi.org/10.1186/2046-4053-4-6>
- Ring, N., Ritchie, K., Mandava, L., & Jepson, R. (2011). *A guide to synthesising qualitative research for researchers undertaking health technology assessments and systematic reviews*. Retrieved from http://www.healthcareimprovementscotland.org/our_work/technologies_and_medicines/programme_resources/synthesising_research.aspx
- Robertson-Malt, S. (2014). Presenting and interpreting findings. *American Journal of Nursing*, *114*, 49–54. doi:10.1097/01.NAJ.0000453044.01124.59
- Rojon, C., & Saunders, M. N. K. (2012). Formulating a convincing rationale for a research study. *Coaching: An International Journal of Theory, Research and Practice*, *5*, 1–7. Retrieved from <http://doi.org/10.1080/17521882.2011.648335>
- Saini, M., & Shlonsky, A. (2012). *Systematic synthesis of qualitative research*. New York, NY: Oxford University Press.
- Sandelowski, M., & Barroso, J. (2002). Reading qualitative studies. *International Journal of Qualitative Methods*, *1*, 74–108.

- Retrieved from <https://journals.library.ualberta.ca/ijqm/index.php/IJQM/article/view/4615/3764>
- Shaw, R. L., Booth, A., Sutton, A. J., Miller, T., Smith, J. A., Young, B., . . . Dixon-Woods, M. (2004). Finding qualitative research: An evaluation of search strategies. *BMC Medical Research Methodology*, 4, 1–5. Retrieved from <http://doi.org/10.1186/1471-2288-4-5>
- Soilemezi, D., Drahota, A., Crossland, J., & Stores, R. (2017). The role of the home environment in dementia care and support: Systematic review of qualitative research. *Dementia*. Retrieved from <http://doi.org/10.1177/1471301217692130>
- Spencer, L., Ritchie, J., Lewis, J., & Dillon, L. (2003). *Quality in qualitative evaluation: A framework for assessing research evidence*. London, England: Cabinet Office, Government Chief Social Researcher's Office.
- Stern, C., Jordan, Z., & McArthur, A. (2014). Developing the review question and inclusion criteria. *American Journal of Nursing*, 114, 53–56.
- Toews, I., Booth, A., Berg, R. C., Lewin, S., Glenton, C., Munthe-Kaas, H. M., . . . Meerpohl, J. J. (2017). Further exploration of dissemination bias in qualitative research required to facilitate assessment within qualitative evidence syntheses. *Journal of Clinical Epidemiology*, 88, 133–139. doi:10.1016/j.jclinepi.2017.04.010
- Tong, A., Flemming, K., McInnes, E., Oliver, S., & Craig, J. (2012). Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Medical Research Methodology*, 12, 181. Retrieved from <http://doi.org/10.1186/1471-2288-12-181>
- Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care: Journal of the International Society for Quality in Health Care/ISQua*, 19, 349–357. Retrieved from <http://doi.org/10.1093/intqhc/mzm042>
- Tsafnat, G., Glasziou, P., Choong, M. K., Dunn, A., Galgani, F., & Coiera, E. (2014). Systematic review automation technologies. *Systematic Reviews*, 3, 74. Retrieved from <http://doi.org/10.1186/2046-4053-3-74>
- Wilczynski, N. L., Marks, S., & Haynes, R. B. (2007). Search strategies for identifying qualitative studies in CINAHL. *Qualitative Health Research*, 17, 705–710. Retrieved from <http://doi.org/10.1177/1049732306294515>