Chapter 6: Selecting studies and sampling

Heather Ames, Andrew Booth, Jane Noyes

To cite this chapter, please use: Ames, H, Booth A, Noyes J. Chapter 6: Selecting studies and sampling. Draft version (May 2024) for inclusion in: Noyes J, Harden A, editor(s). *Cochrane-Campbell Handbook for Qualitative Evidence Synthesis*, Version 1. London: Cochrane

Key points:

- Study selection (also called "sifting" or "screening") is the process of selecting studies that meet the inclusion criteria for a qualitative evidence synthesis (QES) in a systematic and transparent way.
- Study selection is important as it forms the foundation for the data to be subsequently synthesised.
- Study selection in a QES can be comprehensive (trying to find everything) or purposive (trying to find enough to represent a diverse set of experiences and contexts).
- Sampling in a QES can occur when too many studies meet the inclusion criteria for a QES.
- Study selection and sampling in a QES is often iterative with decisions and methods evolving as the review progresses.
- Software and machine learning functions are available to support the study selection process.
- Clear and transparent reporting of the study selection and sampling process is required as part of the audit and decision trail.

6.1 INTRODUCTION

Once review authors have completed a qualitative evidence synthesis (QES) protocol and conducted one or more searches, they can turn their attention to study selection. Study selection involves ruling in studies that meet the inclusion criteria and address the review question and ruling out studies that do not.

This chapter is important as methods for selecting studies for a QES or a mixed-methods review with a qualitative component have evolved in recent years and there is little up to date guidance for review authors to follow. Initially, many QES methods carried an implicit assumption that, as a type of systematic review, the sample would be comprehensive and include all relevant studies. Guidance on the conduct of critical interpretive synthesis (see chapter 19) was probably the first to actively endorse a more flexible approach to sampling qualitative studies (Dixon-Woods, Cavers et al. 2006). Information presented in this chapter will help fill this gap and provide guidance on study selection and sampling for a QES.

Review authors should also read chapter 13 on the GRADE CERQual approach for assessing confidence in synthesised qualitative findings. The relevance, adequacy, coherence and methodological limitations components of GRADE-CERQual all have implications for study selection and sampling. The GRADE-CERQual adequacy component in particular takes account of richness of the studies contributing to a synthesised finding. The adequacy of data in the QES overall and at the level of synthesised findings can be influenced by study inclusion and selection decisions. The GRADE-CERQual relevance component also looks at the fit between the context specified in the review protocol and the context reported in the primary studies. Assessing this contextual fit at individual study level is an integral part of the study selection and sampling process that can impact on the subsequent GRADE-CERQual assessments of relevance at the level of synthesised findings. If overall GRADE-CERQual assessments for specific findings fall below high or moderate confidence, review authors may need to take an iterative and flexible approach moving backwards and forwards between GRADE-CERQual assessments conducted after findings have been developed and the original sampling decisions to identify the studies for inclusion in the synthesis to see if the inclusion of additional studies could further strengthen overall CERQual assessments. Chapter 7 also provides further guidance on assessing methodological strengths and limitations in primary qualitative studies that can be used in conjunction with study sampling.

The chapter begins by providing guidance on how to select studies for a QES, or mixed-methods review with a qualitative component and when to use different types of study selection approaches. A summary of the differences between comprehensive and purposive study selection and sampling is provided, and the typical processes and decisions for selecting studies are then outlined. An overview of which software can be used to aid study selection is presented. The chapter will then cover what to do if the search identifies too few or too many studies, how to sample studies for inclusion in a synthesis including using a tool to assess conceptual richness and contextual thickness, and how to report the

study selection and sampling process. Finally, the chapter concludes with sections on stakeholder engagement and involvement, review author reflexivity, and equity, diversity and inclusion in relation to study selection and sampling.

6.2 SELECTING STUDIES FOR INCLUSION

Box 1 outlines a set of expectations for review authors concerning study selection and sampling. The guidance in the following sections shows review authors how to meet these expectations.

Box 1. Expectations of review authors concerning study selection and sampling:

1) Ensure screening and study selection is **congruent** with the synthesis, both methodologically and practically.
2) Report screening, study selection and sampling transparently. This includes not only the procedures (the **audit trail**) but also the rationale behind decisions and a clear explanation of choices made (the **decision trail**).
3) Include a **PRISMA diagram** demonstrating study selection.
4) Provide a clear **list** of:
    a. **Excluded studies** (studies that did not meet the inclusion criteria)
    b. **Included studies** (all studies that met the inclusion criteria)
    c. If you sampled: **Sampled studies** (studies included in the synthesis)
5) Demonstrate that sampling for a QES is often **iterative** and **dynamic**.

The study selection and sampling process in a QES, or within the qualitative component of a mixed-methods review often takes an iterative approach where decisions and methods evolve as the review progresses. Iteration and flexibility mirrors the approach taken in primary qualitative research where research questions and data collection methods can change and evolve over time. As outlined in Chapter 2 on question formulation, initial scoping searches and stakeholder engagement and involvement can be used to develop the review question and scope and lead to refining or expanding the topic, phenomena of interest and inclusion criteria, which helps to refine the study selection process. Common to all review types, too little or too many studies can also lead review authors to adjust the review scope or research question to conduct a good quality synthesis. An alternative strategy if too many studies are identified is to sample.

It is generally not considered essential to identify and include every available relevant study in a QES (See Chapter 13 on GRADE-CERQual for further definitions on the different types of relevant studies). The most important consideration is to include relevant studies that

represent, for example, diverse participants, contexts, theoretical concepts, research questions and phenomena of interest outlined in the review protocol (Noyes, Booth et al. 2018, Downe, Finlayson et al. 2019). Therefore, study selection in a QES can be complex and sometimes complicated as the review authors need to decide whether to include all studies (comprehensive approach) or to select a sample of studies (purposive approach) or both (Noyes, Booth et al. 2019)(see also section 6.2.4).

Study selection (also called "sifting" or "screening") is critical to the review process. It is part of ensuring that synthesised findings presented in the QES are trustworthy, rigorous and useful; it is therefore important to avoid study selection that is idiosyncratic and unplanned (Popay, Mallinson et al. 2010). Key to study selection are clear eligibility criteria; to make clear which studies are to be included and which studies are to be excluded, and to limit the number of titles and abstracts that fall into an intervening 'corridor of uncertainty'. A 'corridor of uncertainty' refers to studies that purport to meet the inclusion criteria, but on closer inspection they appear to have little relevance or value to the synthesis. This separation is particularly challenging when titles use metaphors or quotations, or abstracts are not structured, which can be common in publications reporting qualitative research (Booth 2016).

The number of studies included in a QES needs to be manageable for the review authors to conduct a high-quality synthesis. It is critical that a transparently-reported audit trail is maintained for study selection and sampling so that others could reproduce the methods and processes if necessary (Porritt, Gomersall et al. 2014). Transparency is the guiding principle when reporting all decisions and their rationale in an audit and decision trail, as review authors may select a varying sample of studies even when addressing similar questions (Noyes, Booth et al. 2019).

6.2.1 Studies versus reports/publications as the unit of interest

A search for primary qualitative studies may identify multiple reports/publications from the same qualitative study. The study becomes the unit of interest while all reports/publications may meet the inclusion criteria and contribute data for the review. For example, one study may comprise three separate relevant reports/publications that are included in the review. The PRISMA flow diagram (see figure 4) should document a single study, adding an annotation that this constitutes three separate reports/publications.

6.2.2 Trial sibling and unrelated studies

It is relatively common for multiple reports/publications ("sibling studies") to be associated with a primary quantitative study such as a trial (Noyes, Hendry et al. 2016). Chapter 5 provides guidance on searching strategies and methods to identify multiple reports from the same study.

One of the decisions that needs to be made when selecting studies for a QES that is linked to a review of intervention effect is whether to include trial sibling and/or unrelated

qualitative studies. A qualitative trial sibling study is a qualitative study that is carried out in conjunction with or during a trial (Noyes, Hendry et al. 2016). An unrelated study is a qualitative study that is carried out independently and not linked to a trial. As a general rule, review authors are encouraged to search for both trial sibling and non-trial sibling qualitative studies for inclusion in their review as the synthesis will likely be better quality if based on data that covers all perspectives and phenomena of interest outlined in the protocol. Some QESs, are however designed to explore why specific interventions do or do not work and may privilege including trial sibling studies because trial sibling studies that collect data from the same participants in the included trials have high contextual relevance. Whereas qualitative studies that address relevant issues about a comparable intervention in a comparable context but are unrelated to a trial are considered conceptually relevant and provide additional enriched perspectives (Noyes, Hendry et al. 2016, Noyes, Booth et al. 2019). Combining sibling and non-sibling studies also provides a richer and larger dataset across multiple contexts to better understand the phenomenon of interest (Noyes, Hendry et al. 2016, Noyes, Booth et al. 2019) .

Whether trials include qualitative sibling studies or not may be topic or intervention dependent. One hundred trials from the register of the former Cochrane Effective Practice and Organisation of Care Group were only associated with thirty qualitative studies (Lewin, Glenton et al. 2009). In contrast, another QES about directly observed treatment for TB found five out of six trials in a Cochrane review were linked to qualitative research, although not all had been published and two studies required translation (Noyes and Popay 2007). A detailed analysis found that both sibling and non-sibling qualitative studies can make a useful contribution to the synthesis (Noyes, Hendry et al. 2016).

In some instances, authors who conduct a QES in conjunction with a review of interventions may decide to only include qualitative trial sibling studies, thereby limiting the potential sample for synthesis. For example, a mixed-methods review for a rapid health technology assessment to explore the experience of exercise referral schemes only included qualitative studies if they had been conducted as sibling studies alongside an RCT (Campbell, Holmes et al. 2015). Another approach is to identify and include trial sibling studies along with non-sibling studies and to undertake a subgroup analysis to see if the non-sibling studies provide unique insights and findings. Only including trial sibling studies will likely have implications when conducting GRADE-CERQual assessments as there may be fewer included studies (adequacy), variation in study quality (methodological limitations), but the trial sibling studies will be highly contextually relevant to the question (relevance). Generally, review authors may also want to consider conducting additional targeted searches to address any gaps if only including trial sibling studies. Chapter 5 provides guidance on when to update the search based on GRADE CERQual assessments of confidence in the synthesized findings (See Chapter 13 on Grade CERQual).

6.2.3 Language considerations

When developing and refining the study eligibility criteria for a QES, review authors need to decide how to handle studies published in different languages. Due to the time consuming, nuanced, and complicated nature of translating and interpreting qualitative data, review authors often decide only to include studies in languages spoken fluently by review authors. If the decision is to include studies in any language, then review authors typically identify such studies during title and abstract screening. Where the abstract is written in a language outside those spoken by review authors, translation software such as Google Translate (Google 2021), Chat GPT or DeepL Translator (DeepL 2021) can be used to determine if the abstract meets the inclusion criteria.

The same translation software can be used to establish the eligibility of studies that proceed to full text screening. Although such translation may not be specific enough to be used during data extraction and synthesis it is typically sufficient to determine if the article meets the inclusion criteria.

6.2.4 Types of sampling methods

Each QES will require a bespoke sampling strategy. Review authors face a choice between comprehensive study selection (trying to find everything) and or a purposive approach (trying to find enough to represent a spectrum of experiences, opinions and aspects of context specified in the protocol). It is also common for more than one sampling strategy to be used in a QES. For example, most QESs start off with a comprehensive search to identify all studies and then an iterative decision can be made regarding the need for a subsequent stage(s) of purposive sampling.

Purposive sampling also includes many different sub-types that have similarities and key differences. One of the methodological limitations of listing the purposive sample subtypes in a table with a single published example of its application in a QES is the danger that the different sub-types appear to be fixed and inflexible when the boundaries between them are not fixed and there is often overlap as well as difference. With the aforementioned caution, readers can refer to Table 1 in Benoot et al 2016 (Benoot, Hannes et al. 2016) or Suri 2011 (Suri 2011), which provides a summary of purposive sampling approaches that have been used in different primary qualitative research studies and QESs. See also the additional file on the handbook website.

Selecting the most appropriate purposive sampling method(s) is therefore a key decision (or a number of iteratively made decisions) in a QES and a decision that ideally needs to be informed by an experienced qualitative researcher/reviewer and a reviewer or stakeholders (including patients and the public) with topic specific knowledge. When, where and which purposive sampling method(s) to use may evolve as the review progresses and is another example of the iterative way that QES protocols and reviews can evolve over time. Many review authors also use more than one purposive sampling method in their QES. For example, Benoot and colleagues used three types of purposive sampling in their QES. They

started off with an intensity sample to select data rich studies that did not report very unusual cases. Then they constructed a maximum variation sample to select studies that included key dimensions of interest and wide variations amongst cases. Then they undertook disconfirming case sampling to identify studies that did not fit the emerging synthesis and their interpretation to provide rival interpretations as a way of creating a boundary around their core synthesised qualitative findings.

In the next section a typical process is outlined for selecting studies. As appropriate, review authors can build on this typical process to create a bespoke process for their individual QES. If review authors wanting to purposively sample but are not sure of where to start, in section 6.5.1 below, guidance on getting started with a maximum variation sampling frame is described and an example is provided as to how Ames developed a maximum variation sampling framework to purposively sample studies in a Cochrane QES. In section 6.5.2 guidance is also provided on how to use a data richness/thickness tool to support sampling decisions.

6.2.5 A typical process for selecting studies

Figure 1 illustrates a typical process for study selection and the choices to be made. Diamonds represent choices the review authors need to make. Several considerations help when deciding how to proceed at each stage. Start with the RETREAT framework: Review question, Epistemology, Time/Timescale, Resources, Expertise, Audience and purpose, Type of data outlined in Chapter 8 on choosing a method of synthesis. Review authors should think about how each criterion might inform their choice of screening process; what do the stakeholders and review commissioner want? will the QES enlighten understanding or inform guidelines? Review authors can talk with patient and public representatives, other key stakeholder groups and the funder to ascertain their perspectives on what is important for the specific review. Importantly, review authors need to be realistic and aware of what is feasible and desirable with the available pool of data. Early scoping searches can also help to determine the size and potential quality and richness of available evidence (See Chapter 5 on Searching). If searches yield too many studies, then review authors may consider reducing the timeframe of publication to reduce the number of potential studies or opt for sampling. The protocol will need updating iteratively as and when decisions are made. Each decision needs to be taken on a review-by-review basis and a clear audit trail needs to be documented with a clear rationale provided for each decision made (decision trail).

Typically, the study selection process begins with a comprehensive search and deduplication of the identified references. After this the review authors begins to screen or sift through the studies identified in the search, making decisions on the title/abstract first and then on the full text of papers that appear to meet the inclusion criteria. Once review authors have completed screening, they can assess whether they have too little data, an appropriate amount of data or too much data and proceed accordingly to decide which

studies will be included in the synthesis. At this stage a decision also needs to be made as to which method of synthesis is the most appropriate for the type and amount of data (Figure 1 and Chapter 8 on selecting a method of synthesis).

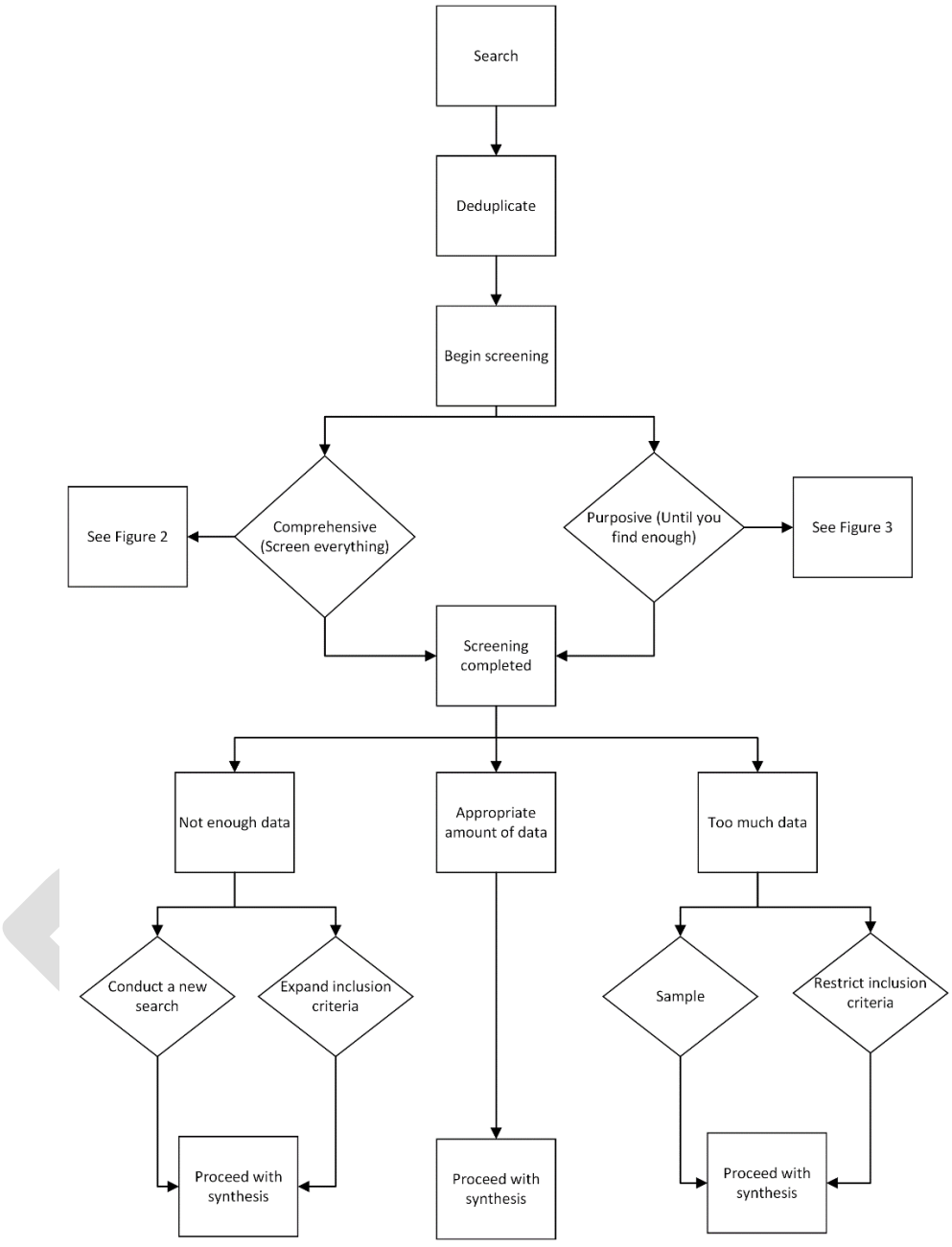


Figure 1: The study selection process

A comprehensive study selection approach (figure 2) follows a traditional systematic review approach. First, screen all titles and abstracts. Second, screen full texts. Third, make decisions around the adequacy of data and how to handle adequacy once all eligible studies are identified. An extra decision loop is shown in Figure 2 to represent common variations in screening and study selection in QES. First, it is common in a QES for the inclusion criteria to be adjusted as the review proceeds. A loop during screening is included to represent this

stage. Review authors may choose to revisit their sampling and synthesis plan after having mapped the studies that meet the inclusion criteria by conducting a data extraction of study characteristics. Many review authors find it helpful to produce a map or excel spreadsheet to display studies and their characteristics that meet the inclusion criteria.

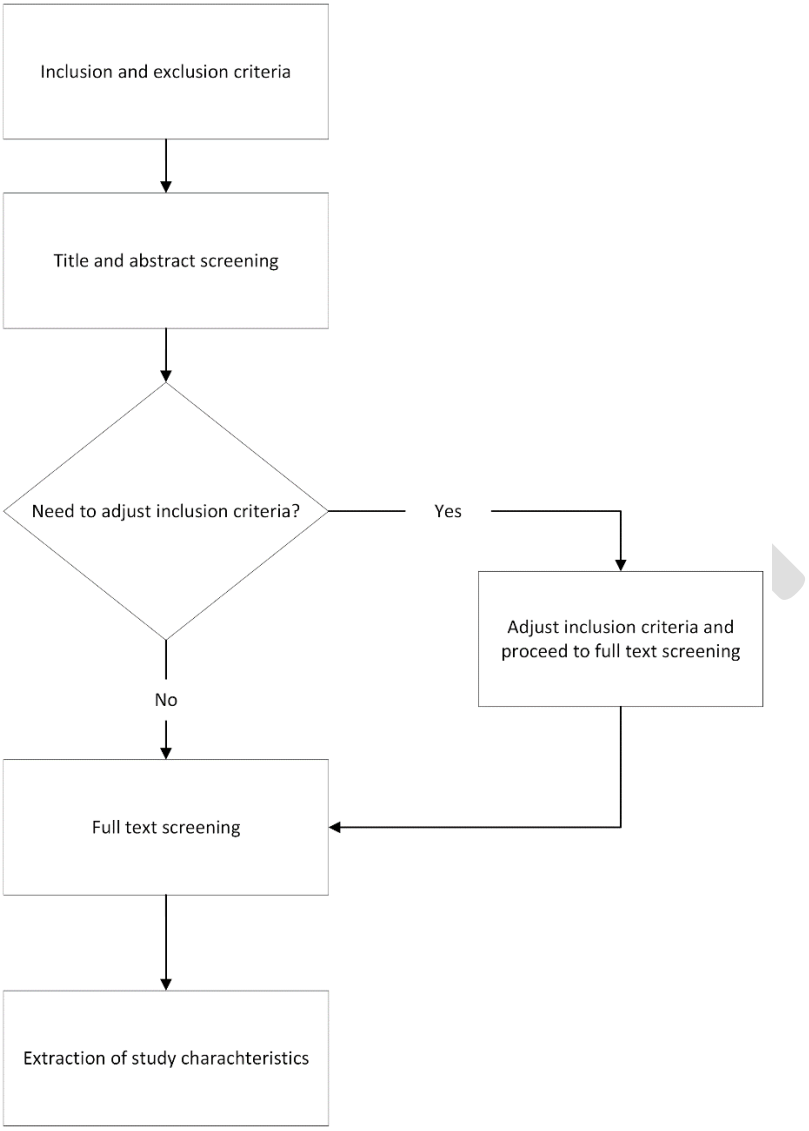


Figure 2: A comprehensive study selection strategy

In a purposive approach to study selection (Figure 3), the process is iterative and evaluation ongoing. There is no formula for deciding when sufficient studies have been included in the synthesis. The decision about when to stop sampling studies needs to be made on a review by review basis and will depend on a multitude of factors, such as:

- the selected method of synthesis and its capacity to include studies,
- the sub-type of the purposive sampling method(s) selected,
- the richness/thickness of the data,
- methodological limitations in eligible studies,

- wanting to include studies representing all aspects of the context specified in the protocol,
- the question being asked,
- the amount and type of identified studies from which to sample,
- subsequent GRADE-CERQual assessments of synthesized findings

Review authors may decide to stop screening once they feel they have identified sufficient studies that meet the inclusion criteria and provide rich/thick data to sustain a synthesis. For example, review authors could search in one database and screen these studies first or choose to screen the most recent studies first as they are the most contemporary to current experience and practice. While screening the full text of studies that appear to meet their inclusion criteria review authors can further discuss and clarify concepts and participants and decide to adjust their inclusion criteria before searching further. Review authors could also choose to extract data from these studies and begin their synthesis, using the next round of searching to build on emerging ideas or patterns in the data. If, in the end, review authors find that the amount of data is too large to enable a quality synthesis they may choose to sample from the studies that meet their inclusion criteria. In addition, review authors can decide to stop sampling at a key tipping point when the addition of another study does not change the findings. If using a key tipping point to stop sampling, then review authors should start by including the most up to date and contemporary evidence first and work backwards along the timeline for study inclusion.

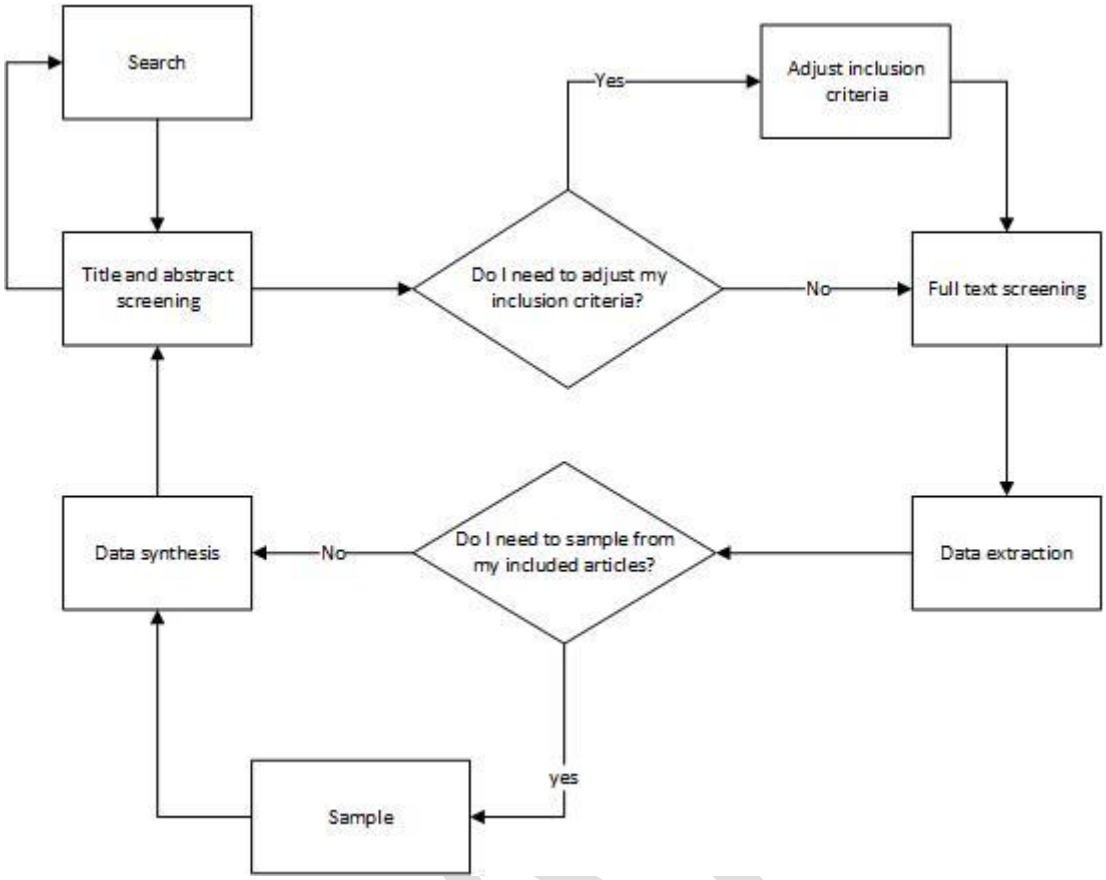


Figure 3: An example of a purposive approach to study selection

6.2.6 Applying study selection and eligibility criteria

Although study selection typically involves sifting through numerous titles and abstracts, eligibility criteria are applied at four points:

(i) during the literature search (for example limits by date, language or study or publication type (Chapter 5)

(ii) screening titles and abstracts, typically at between 60 and 200 abstracts an hour (Edwards, Clarke et al. 2002, Shemilt, Khan et al. 2016)

(iii) scanning the full texts for the presence or absence of key details; and in some cases

(iv) assessing the studies for design features, methodological limitations (Chapter 7), or data thickness/richness – see section 6.5.2.

Study selection usually takes place following completion of topic-based searches and supplementary searching (chapter 5). By this time review authors have usually formed a good idea, based on the numbers and types of search results, as to whether eligibility criteria should be ‘strict’ or ‘forgiving’. For example, should particular inclusion criteria be fulfilled by a mention in the title and abstract or is it sufficient for a topic of interest to be present in a section of a full-text paper? Forgiving criteria may require considerable full-text reading and be time consuming. Strict criteria may risk an ‘empty review’ or, at the very least, inadequate data to support each finding.

Put broadly the focus of the title and abstract stage is to *rule out* obviously irrelevant studies. Studies that are unclear are carried over to the full text stage along with all eligible studies. In contrast, the full-text stage seeks to *rule in* a final set of relevant studies. No studies should be unclear beyond this point.

Review authors are advised not to select studies 'live' from a bibliographic database, as this may hinder transparency in reporting the study selection process. The review authors may fail to document their verdicts and the session may be interrupted. Instead, studies should be downloaded from each database and imported into reference management software (e.g., EndNote) for selection and/or imported into systematic review software (e.g., EPPI Reviewer (www.eppi.ioe.ac.uk), Rayyan (www.rayan.ai), Covidence (www.covidence.org). Some review authors use reference management software for study selection, either coding each individual item or dragging the record into an Include, Exclude or Query folder. Others export data from their reference management software into a spreadsheet and then use dropdown menus to select verdicts for each criterion. However, with large search result sets this type of screening can be time consuming. An increasingly popular, and recommended, option is to download studies and import them into study selection review software such as Covidence (www.covidence.org), EPPI Reviewer (www.eppi.ioe.ac.uk), CADIMA (www.cadima.info) or Rayyan (www.rayan.ai) (see section 6.3.1). This also allows for easier reporting and auditing of, and transparency in, the screening process.

By the completion of the QES protocol and search, review authors have a basis for piloting the eligibility criteria, to assess consistent understanding of inclusion criteria, identify and resolve ambiguities, further clarify inclusion criteria if necessary, and add to a screening checklist or instructions for review authors conducting the screening. Piloting should be conducted by at least two review authors to identify any potential ambiguities, but ideally by everyone who is to be involved in study selection. When identifying the criteria, it is helpful to consider two issues; (i) at which stage is each criterion best determined? At title/abstract or full text? And (ii) in which order should the criteria be considered? For many topics it is easiest to rule abstracts out based on study design (Does this study report qualitative research?). However, for other topics it is quicker to rule in or out based on population (Does the study involve children/adolescents?) or context (Was the study conducted in a school?).

During this meeting, review authors can consider making and using a screening checklist to follow with 'yes' or 'no' answers. A 'yes' answer means proceed and a 'no' answer means exclude. This can be used for both title/abstract and full text screening. For example:

1- Was the study published within the dates for inclusion?
2- Does the study meet the language inclusion criteria?
3- Does the study use qualitative methods for data collection and analysis?
4- Does the study explore the appropriate population?
5- Does the study explore the phenomenon of interest?

At this meeting, review authors should set aside time to screen items together so that discrepancies and grey areas can be discussed as they arise.

The gold standard within systematic reviews of effectiveness requires double blind screening with consensus reached on conflicting decisions. Screening for a QES or mixed-methods review with a qualitative component can use similar approaches. Alternatively, review authors can adopt an open consensus-based approach given that decisions about study inclusion may involve a wider discussion with key stakeholders, patient and public representatives or the funder. Some review authors may also undertake double blind screening of titles and abstracts with all conflicting decisions to include/exclude being resolved by a third person or by consensus agreement at the end of title and abstract screening, or taken forward to full text screening for resolution. This allows for capturing different perspectives and interpretations of a non-objective phenomenon (the topic of interest for example) to offer greater diversity amongst the included studies. The review authors could then discuss interpretation of inclusion criteria at the full text screening stage. Involving two reviewers in as much of the screening as resources permit allows for consensus but may usefully reveal different perspectives on the focus of interest (Booth, Carroll et al. 2013). For an overview of screening approaches see Table 1.

In reviews that use machine learning ranking algorithms to quickly identify relevant studies, or reviews with very precise small searches, review authors should sit together to screen to reach clear agreement on inclusion criteria. Review authors can then discuss references that lie within the grey area of their inclusion criteria. When screening extensive search results, where the aim is for speedy removal of all studies that clearly do not meet the inclusion criteria, single screening can remove these studies without impairing the screening process. In this situation review authors should check in regularly with each other to determine when to switch to double screening.

In some cases, studies can be distributed among individual review authors trusting each to independently make a definitive verdict (single screening). Although not recommended practice for a Cochrane review, time limits or capacity restrictions may necessitate single screening (e.g. in a time sensitive review – see chapter 15). If this is the case, review authors should agree on a detailed inclusion checklist. They should sit and screen together until a consensus on the inclusion criteria is reached. After this, review authors can screen independently with a very low threshold for contacting their co review author with questions on studies where they are unsure.

Table 1: An overview of screening approaches

First screener	Second screener	Piloting	Check sample	Consensus required?	Notes
Random double-blind screening of all titles/abstracts		Minimum 5% of all records	10-20% of all records (including pilot)	Yes - on conflicts	Gold standard for intervention reviews
Random double-blind screening of all titles/abstracts		Minimum 5% of all records	10-20% of all records (including pilot)	No - all conflicts rolled over to full-text	Allows for diverse perspectives
Double screening of high-relevance titles/abstracts; single screening of remainder		Joint screening to determine criteria	Predetermined relevance level on algorithm	Yes - On conflicts or No- roll over to full-text	Used with machine algorithms or small search sets
Screening 50% of titles/ abstracts for exclusion	Screening 50% of titles/ abstracts for exclusion	Minimum 5% of all records	Double screening of includes	Yes - On conflicts or No- roll over to full-text	Minimises risk of false negatives
Screening 50% of titles/ abstracts	Screening 50% of titles/ abstracts	Predetermined number (not less than 5%) or level of consensus	10-20% of all records (including pilot)	No consensus	Risk of false negatives
Single screening of all titles/abstracts		Minimum 5% of all records	No check	No consensus	Not recommended

Once a study has been determined to meet the inclusion criteria from the title and abstract, review authors should seek to obtain a full-text copy of the reference. Full text retrieval can be achieved through publication identification features of review or reference management software, or through using application program interfaces (APIs) to communicate with publishers. Full text screening can be conducted simultaneously with title and abstract screening. This prevents unnecessary delays and allows the authors to start designing a data extraction form based on clear Includes. Selecting full-text studies for final inclusion requires an unambiguous verdict (satisfying all inclusion criteria and no exclusion criteria). No “unclear” verdicts should persist beyond this full-text stage; such verdicts either mean that the eligibility criteria are poorly designed or suggest a need to contact study authors for missing details.

6.3 SOFTWARE SUPPORT FOR SELECTING STUDIES

A search for eligible studies can often identify thousands of records to be screened. Selecting studies from among these records can be particularly time-consuming and logistically challenging. These and other challenges have led to the development of software tools and packages that offer support for study selection and help authors to maintain a clear audit trail and facilitate reporting.

Software to support selecting studies can be classified as:

- systems that support study selection, typically involving multiple reviewers (see Section 6.3.1); and
- tools and techniques based on text mining and/or machine learning, which aim to semi- or fully-automate study selection (see Section 6.3.2).

6.3.1 Software for managing study selection

Managing study selection and maintaining a clear audit trail is challenging, particularly in QESs that involve multiple reviewers and that can take an iterative approach to study selection. Numerous software programs are available to support study selection. So far, no software supports all steps of each type of QES. However, such tools are useful during screening and study selection. Software to support study selection, along with other stages of a systematic review, including text mining tools, can be identified using the Systematic Review (SR) Toolbox. The SR Toolbox is a community driven, web-based catalogue of tools that support systematic reviews (Marshall and Brereton 2015).

Different stages of study selection can be automated including full-text retrieval, de-duplication, and dividing studies into groups for screening, using reference management software (EndNote, Reference Manager), generic software (such as Microsoft Excel), or purpose-specific review software (EPPI Reviewer, Rayyan, CADIMA, Covidence).

6.3.2 Automation and machine learning during study selection

The decision on whether to include or exclude a study can also be automated through machine learning (EPPI Reviewer, Rayyan, Covidence). Research into machine learning has received considerable attention resulting in development of various tools and techniques (Higgins, Thomas et al. 2019). Most research on the use of machine learning during study selection has been conducted for systematic reviews of effectiveness (O'Mara-Eves, Thomas et al. 2015, Bannach-Brown, Przybyła et al. 2019, Gates, Guitard et al. 2019, Callaghan and Müller-Hansen 2020). However, machine learning holds significant potential to reduce the study selection workload for all types of reviews (Thomas, Noel-Storr et al. 2017). Research suggests that using a ranking algorithm can reduce manual screening by between 30% and 70%, although sometimes at the cost of a 5% reduction in sensitivity (O'Mara-Eves, Thomas et al. 2015).

Where two reviewers are screening each study, machine learning can be used to automate one or both reviewer author's decisions. Machine learning in study selection typically refers

to models learning from humans *include* or *exclude* binary decisions, then applying that learning to unscreened studies. Suggested inclusion decisions can be presented in binary form as a classification prediction, and authors may then evaluate the prediction and accept or reject it. A study design classifier that can predict whether a study is a *qualitative study* or *not qualitative study* is in pressing need of development.

A ranking algorithm can generate a continuously sorted list according to the review author's screening decision, rather than binary (In/Out) prediction. Ranking algorithms regularly learn as review authors screen, and sort and re-sort studies into a list in which likely included studies are presented first (ranked) (O'Mara-Eves, Thomas et al. 2015). With a successful model, review authors can speed up the process of identifying most of the relevant studies without having to replace human screening. When using a ranking algorithm, the screening process becomes more intensive as relevant references are pushed forward. It is important for review authors to screen together and have good discussions to clarify inclusion criteria early on. It is also important to resolve screening conflicts more frequently in order for the algorithm to update based on new decisions.

Review authors can decide at which point to stop screening and exclude the remaining, highly irrelevant studies without manual assessment (i.e., "automatic exclusion"). So far, no gold standard method exists to pre-specify the optimal point at which to stop screening. Numerous criteria are suggested and the cut-off point depends on an acceptable level of risk, or the trade-off between precision and recall (Callaghan and Müller-Hansen 2020). See also section 6.5.3.

The usefulness of ranking algorithms extends beyond automatic exclusion, since by prioritizing records by relevance, authors identify studies for inclusion earlier than otherwise possible. Importantly, such identification expedites piloting, as authors typically can focus on grey areas relating to relevant and potentially relevant studies rather than clearly irrelevant studies.

6.4 STEPS ONCE INCLUDED STUDIES ARE IDENTIFIED

In a QES too much data can threaten the quality of the synthesis by rendering the synthesis unmanageable (Sandelowski 1995, Morse 2010). It can be caused by the inclusion of too many studies, the inclusion of too much data in the included studies or by a lack of review author experience during scoping (looking through the studies in preliminary searches or looking through those that meet inclusion criteria). Conversely, a search may identify very few studies that meet the inclusion criteria which can prove equally problematic but for different reasons.

Ideally an initial scoping of the studies that meet inclusion criteria provides an approximation of the likely numbers of studies and an initial assessment of whether they contain thick or thin contextual detail. However, underestimates and overestimates of potentially relevant studies are not uncommon. Overestimation may require a modified search strategy; to be agreed by the entire authors and tracked via the audit trail.

Comparison revealed that seven times as many studies were retrieved by country-specific literature searches as were retrieved for the same countries within multi-context searches (Booth, Mshelia et al. 2019). However, trading off breadth for depth is usually manageable within existing resources – the overall number of studies is not too great to subvert comprehensive sampling. In contrast, review authors may assume that they will identify too many studies and then revise their strategy if the total number does not turn out to be prohibitive. For example, one Cochrane review author team initially anticipated having to sample from their eligible qualitative studies (Pollock, Campbell et al. 2020). Subsequently, the review authors assessed whether the included studies were sufficiently numerous or rich in data to require a sample of studies (Cochrane Effective Practice and Organisation of Care (EPOC) 2017). They concluded that "due to the relatively low number of included studies, discussion amongst review authors... led to the decision not to select a sample of studies, but instead to extract data from all included studies"(Pollock, Campbell et al. 2020)(page 15). Several commentators (Thorne 2017, Bergdahl 2019) have also expressed concern at superficial analyses and syntheses resulting from large numbers of conceptually poor studies, time constraints or analytical naivety of the reviewers.

A consensus on the optimal number of studies to include in a QES is unlikely, if not inconceivable. One reason for variation relates to the different data requirements of different types of synthesis methods (Campbell, Pound et al. 2011). Synthesis methods using comprehensive search strategies seek to gather data from all eligible participants or studies ('*every participant/study counts'*) whereas synthesis methods that employ purposive sampling strategies aim to develop a new interpretation of the experiences observed (*what 'counts' or matters*). Pragmatic and methodological factors to consider, when thinking through how many studies to include in a synthesis, broadly map to the RETREAT considerations (Booth, Noyes et al. 2018), including:

- The planned type of analysis (Audience & purpose and Epistemology)
- The contextual thickness and/or conceptual richness of the included studies (Type of Data)
- The authors' individual and collective experience with qualitative synthesis (Expertise)
- The amount of time to complete the review (Time and Resources)
- The breadth of the topic of interest, for example, the need for data diversity across contexts or populations (Research question)

The following sections explore options for too few studies, sufficient studies and too many studies for the planned synthesis.

6.4.1 Identification of too few studies

Occasionally review authors find that very few studies meet the eligibility criteria. Review authors face two main options: (I) redefining the review parameters to admit more studies or open the inclusion of qualitative data from open ended questions in questionnaire

surveys that have been analysed using qualitative methods. Questionnaire surveys are however not qualitative studies and usually contain conceptually poor data with little to no contextual detail.

(ii) establishing a conceptual link to "indirectly relevant" evidence. See Chapter 13 on GRADE CERQual assessment of study relevance which outlines a typology of relevance to classify studies (including indirectly relevant studies) for use in this specific context of having too few studies.

Usually, having too few studies results from insufficient scoping of the topic, overly severe eligibility criteria and/or scope limited by co-terminosity (being conducted in conjunction) with a quantitative review (as in a mixed-methods review).

Several decisions can increase the likelihood of finding sufficient studies (Lins, Hayder-Beichel et al. 2014). First, the eligibility criteria can be expanded beyond qualitative studies that utilise recognised qualitative methods of data collection and data analysis to include process evaluations and mixed-methods studies with a qualitative component. Second, broadening the scope of the review, for example the perspectives, topic of interest and/or setting. Third, study design filters can be removed to include search results for any topically relevant studies. The authors then sift more extensively through studies in search of additional relevant items. Even after these modifications one review author team only identified two relevant items for inclusion (Lins, Hayder-Beichel et al. 2014).

As described above, one possible strategy to overcome the problem of too few studies is to search for comparable indirectly relevant evidence from other sources. For example, a QES of values and preferences of lactating mothers with infants affected by Zika virus was extended conceptually to "comparable conditions" that affect infant feeding and swallowing to provide additional indirectly relevant studies (Carroll, Booth et al. 2020). Findings for mothers of infants with cerebral palsy in low- and middle-income countries were found to inform interventions for mothers of children with Zika-related microcephaly and feeding problems.

When review authors make changes in scope they should provide explicit acknowledgement of all changes, justify those changes and document when and how the scope was changed in an update to the protocol or later on in the review process in a section on protocol deviations. Changes are preferably made early in the review, to avoid having to rerun the searches for studies, and should not be made without detailed knowledge of how changes impact upon review findings.

6.4.2 Identification of sufficient studies

If the review authors feel they have identified a sufficient number of studies to adequately address their review question and objectives, they can proceed with the review process. However, QES authors often go back and forth through processes and the GRADE-CERQual assessment of confidence in qualitative findings can provide an indicator as to whether

additional studies could further strengthen a finding and have greater utility in a decision-making process.

6.4.3 Identification of too many studies

Heavily researched topics, such as vaccines or HIV/AIDS, may be populated by many more qualitative studies than required for a synthesis. One response is to introduce *post hoc* inclusion criteria. For example:

- limit the years of publication for the included studies
- narrow the contextual scope of the review from global to regional or national
- narrow the population of the review

In a review of self-management of medical abortion, the review authors added a post hoc requirement, following review of all abstracts, for "a shift to a less medicalized process [to be] outlined in the introduction". This form of purposive intensity sampling required studies focusing on self-management to be privileged (Wainwright, Colvin et al. 2016).

However, if review authors take this approach, they should discuss implications for the confidence in the findings and the usefulness of the finished review (see chapter 13 on Grade CERQual). Narrowing of scope should be based on a clear purpose and fit with the review question and objectives. A decision to narrow the scope of the review may require the review authors to reformulate the research question and objectives to reflect this.

Review authors of QESs do not exclude studies based on methodological limitations as commonly as review authors of effectiveness reviews because a qualitative study with concerning methodological limitations may yield valuable qualitative insights or reveal the attitudes or experiences of an otherwise underrepresented group (see chapter 7).

6.5 Getting started with purposive sampling

Purposive sampling is one way to manage the amount of data for inclusion in the synthesis. Section 6.2.3 signposted to the large number of available purposive sampling sub-types and highlighted the need for an experienced qualitative researcher/reviewer to inform decision-making about their selection. Experienced review authors are generally more confident about which purposive sampling method(s) to use with which synthesis methods as they become more familiar with the available studies and emerging patterns in the synthesised data. Nonetheless, selection of purposive sampling method(s) can involve an element of trial and error until the right combination of methods are identified for any given QES.

Getting started with purposive sampling can be daunting to inexperienced review authors and a lot of time can be wasted if inappropriate sampling decisions are made. Guidance reported here incorporates principles from previous work on this issue (Cochrane Effective Practice and Organisation of Care (EPOC) 2017). As previously mentioned in section 6.2.3, purposive sampling of studies within a QES engage with the same principles used for sampling for primary qualitative research (Silverman 2013, Patton 2014). If inexperienced

review authors are not sure of which sampling method(s) to select then a reasonable place to start is to sample with a purposive maximum variation sampling frame.

6.5.1 Sampling with a purposive maximum variation sampling frame

This section describes the principles underpinning how to develop a purposive maximum variation sampling frame to sample from studies that meet the inclusion criteria.

When constructing a purposive sampling frame, review authors should identify characteristics or variables that are most important to the QES question and objectives and/or where diversity, for example of populations or settings, are known to exist (to create a maximum variation sample). Studies with conceptually rich or contextually thick data may be particularly important - see section 6.5.2 (Ames, Glenton et al. 2019, Ames, France et al. 2023). Once listed, these characteristics or variables should be rearranged within a prioritized order. This prioritized order determines the maximum variation sampling frame. Start by sampling the studies that meet the highest priority and work down the list. A bespoke sampling frame is required for each QES. For example, Ames and colleagues in their Cochrane QES on parents' and informal caregivers' views and experiences of communication about routine childhood vaccination, constructed their maximum variation sample by identifying three key dimensions of variation, and then finding cases that varied from each other as much as possible along these dimensions. They then created a purposive maximum variation sample of 38 studies using the three dimensions of variation that included (i) all studies from low- and middle-income country settings, as a variation from high-income settings; (ii) all articles scoring highly for data richness (iii) examined any remaining studies for those most closely matching review objectives (Ames, Glenton et al. 2017)

6.5.2 Using a study level data richness/thickness assessment tool

Thicker data provides more detailed descriptions of context and richer data provides the foundation for more detailed and transformed conceptual themes, findings, lines of argument and new theories and theoretical insights. Using a study thickness/richness assessment tool can help review authors sample and select thick and/or rich studies to inform their synthesis. QES methods such as meta-ethnography (chapter 11) that set out to transform qualitative data in order to develop new theory require a sample that includes sufficient conceptually rich and contextually thick studies. Ames and colleagues have developed a tool that aids authors in systematically and transparently assessing the conceptual richness and contextual thickness of the data in the studies that meet the inclusion criteria (Ames, France et al. 2024). Early versions of study thickness/richness assessment tool have been piloted in several Cochrane reviews (Ames, Glenton et al. 2017, Ames, Glenton et al. 2019, Campbell, Coleman-Haynes et al. 2020, Cooper, Schmidt et al. 2021, Brown, Carter et al. 2022, France, Uny et al. 2023, Merner, Schonfeld et al. 2023, Wingfield, Kirubi et al. 2023) and were found to help with: sampling decision-making as well

as further enhancing the overall quality of the synthesis; providing additional information to support other processes such as assessing methodological limitations in qualitative studies (chapter 7), and making subsequent GRADE-CERQual assessments of data adequacy and relevance at the level of synthesized findings (chapter 13).

Ames and colleagues provide more detailed guidance on applying the latest version of the contextual thickness/conceptual richness tool (Ames, France et al. 2024). The following text and diagrams are reproduced with permission by Ames and colleagues.

The data thickness and richness assessment tool uses a sliding visual format to help review authors reflect on and assess the level of richness or thickness in their included studies (Figure 4). The same format can be used for assessing both data richness and data thickness for each study. The assessment tool is meant to be used on data that addresses the review question within the primary study. It is important to remember that the assessment process is just the starting point for further discussion. The discussion and consensus process to obtain the assessments are as important as where the studies are located on the scale.

Before authors use the assessment tool, they need to have an overview of the amount of data that addresses the review question in the included study. One option to gain this overview is for review authors to make a judgement about the amount of data in a study that addresses the review question before they go on to assess data richness and thickness. This helps review authors to become familiar with the included studies before proceeding to other assessments. See Table 2 for an example.

Table 2: Categories to facilitate assessment of the amount of data in a primary study that addresses the review question

Amount of data in the study that addresses the review question
Study includes very little data that address the review question
Study includes some data that address the review question
Study includes a moderate amount of data that address the review question
Study includes a large amount of data that address the review question

After making an assessment of the amount of data in a primary study that addresses the review question, review authors can then decide how they want to proceed with the data thickness/richness assessments. For example, they could begin with the studies with the most data addressing the review question first.

Assessing contextual thickness

The data thickness assessment part of the tool aims to help QES authors assess the contextual thickness of the data in a primary study. Contextual and inter-relational descriptions increase the reader's ability to visualise the study participants, context, the intervention (if appropriate), their interactions with the researcher etc.

Depending on the focus of the qualitative research study and question, review authors can potentially assess the thickness of contextual description related to the health and social context (e.g., relevant policy and legal frameworks, the health, social care, education or other type of system and the research problem why the research is important etc), as well as descriptions of the participants (sample), study setting and intervention, and methods and procedures. When using this tool, review authors should look at the contextual thickness of the study as a whole, i.e., across the full study report. Some contextual detail (such as researcher reflexivity and conflicts of interest), for example, may also be reported in the findings, discussion, declarations and supplemental online files.

Assessing conceptual richness

The data richness assessment part of the tool aims to help QES authors assess the conceptual richness of a primary study. In this context, conceptual richness of a study is defined as the degree of abstraction of analysis and interpretation of the data, or what is commonly referred to as the extent of "data transformation", as well as the degree of interpretation of the subjective meaning of participants (Popay, Rogers et al. 1998, Sandelowski and Barroso 2006). This definition in part draws on the typology of the type and nature of qualitative findings developed by Sandelowski and Barroso (Sandelowski and Barroso 2006). This typology conceives qualitative research findings as being located along a continuous spectrum representing the degree of transformation of the data. At one end of the spectrum are less transformed findings; that is, findings that describe patterns in the data. At the other end of the spectrum are more transformed findings that help to interpret and explain the phenomena of interest. These transformed findings have a high level of abstraction and provide theoretical interpretations or explanations of the patterns in the data. In the middle of the spectrum are findings that do more than simply describe the data but are not yet themselves fully transformed data that provide interpretations or explanations of the phenomena of interest. These findings may explore patterns of association in the data and/or link patterns in the data to key theoretical concepts.

When using this tool, review authors should look at the conceptual richness of the whole study, i.e., across the full study report (i.e., not just in the findings section) including text, tables, infographics, photographs, and other visual ways of presenting conceptual information, (Including supplemental files). See table 3 for the assessment tool.

Table 3: Data thickness/richness assessment tool for content relevant to the review question, objectives and context as specified in the review protocol*.

Assessment criteria
Thickness of contextual data
No or very little description that covers minimal aspects of context.
Mostly thin contextual description that covers few aspects of context.
A mixture of some thicker and some thinner contextual description covering some but not all aspects of context, or all aspects but not in sufficient detail.
Mostly thick or very thick contextual description covering most/all aspects of context.
Richness of conceptual data
No or very little transformation of the data in the creation of the findings, and little or no attempt to interpret or explain patterns. No use of theory or conceptual frameworks in the analysis and very little use of relevant empirical literature. No or very little detail on the interpretation of the subjective meaning of actions and behaviours to participants.
Basic application of a theory or conceptual framework to label, present, or organise portions of the data, develop themes or frame the findings. There is little transformation of the data in the creation of the findings and the findings provide little detail on the interpretation of the subjective meaning of actions and behaviours to participants.
Conceptual/thematic description as above but taking it one step further with a degree of theoretical development rooted in the study findings. There is some transformation of the data in the creation of some of the findings, but not the majority. This transformation is intended to further interpret or explain patterns in some aspects of the data or to link these patterns and meanings to key theoretical concepts. Some detail regarding the subjective meaning of actions and behaviours to participants.
The majority of the findings are based on more extensive transformation of the data but some findings may remain very close to the data. The more transformed findings provide theoretical interpretations or explanations of the patterns in the data. These interpretive explanations offer extensions to or propose a new model, framework, theory or line of argument and attempt to provide integrated explanations of phenomena. Theory or a conceptual framework is integrated throughout the paper. Detailed interpretation of the subjective meaning of actions and behaviours to participants.
* The descriptions and explanations in this table draw on the work of Sandelowski and Barroso (2006)

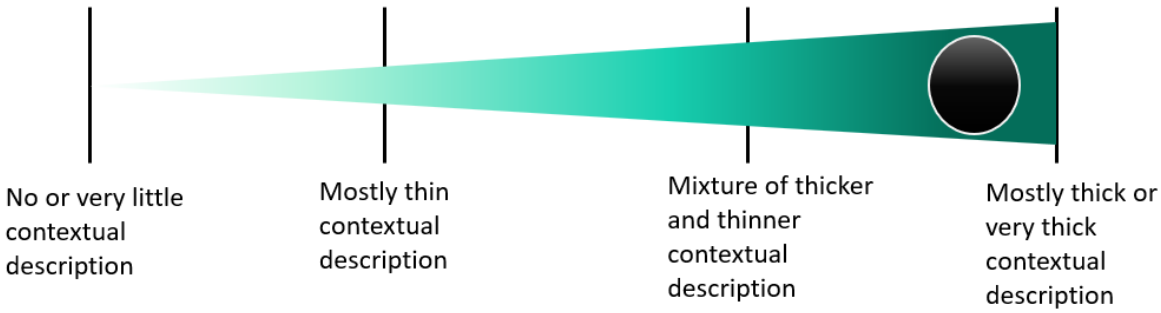


Figure 4: The sliding data (A) thickness assessment tool and (B) richness assessment tool

See Figure 5 for an example of combined thickness and richness assessments.

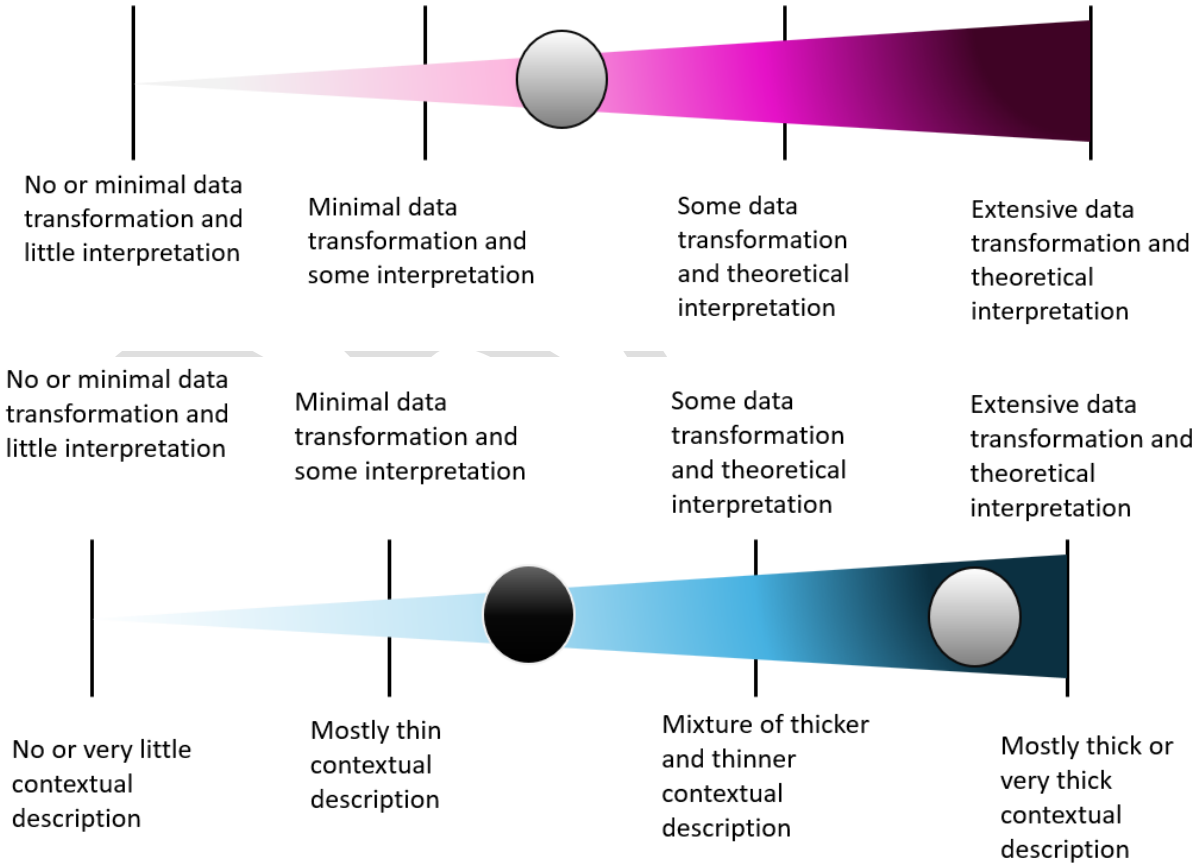
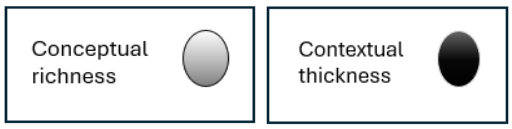


Figure 5: An assessment example where the assessments are recorded on the same tool and different colours are used to differentiate the judgements.



Reporting richness and thickness assessments

Assessments for each study should be reported transparently in an additional file. This could be done visually by assembling all of the assessments on a single (enlarged) tool, creating a colour coded matrix or narratively through description of the review authors' discussions.

6.5.3 When to stop sampling?

No clear guidance or arbitrary number exists for the point at which a review authors can stop sampling studies for inclusion in the synthesis. Stopping depends on the same RETREAT-motivated factors (see section 6.4) such as the richness, thickness, type of relevance of the data, the breadth of contexts and populations, the amount of data that answers the research objective, the type of planned analysis and the experience of the review authors (Booth, Noyes et al. 2018). However, in addition to the RETREAT criteria, other considerations help when deciding whether to synthesise more studies. These include:

- Theoretical saturation - when the authors are confident that are only finding more studies with data leading to similar interpretations (data saturation) (Rohwer, Hendricks et al. 2021). Review authors could continue to sample for dissonance (negative cases) and diversity (Booth, Carroll et al. 2013).
- When the GRADE CERQual assessments of synthesised findings are moderate or high confidence (preferably high confidence).

6.6 REPORTING STUDY SELECTION AND SAMPLING

The study selection and/or sampling strategy and methods (what is planned) must be described transparently in the Methods section. Primary qualitative studies share an almost universal requirement to report their sampling strategy and sample characteristics (who was selected and why) (Gentles, Charles et al. 2016). In contrast, reporting of sampling strategies and sample characteristics (which studies were selected and why) for QES is a relatively neglected topic (See also chapter 20). Review authors should transparently report their study selection and sampling in the Results section. Review authors should leave an 'interpretive trail' (also called an audit or decision trail) to describe how and why studies have been used or omitted (Pawson, Greenhalgh et al. 2005) (Suri 2011), including:

• Describing the sampling logic as matched to the synthesis purpose.
• Describing what logic determined when to discontinue searching.
• Offering justification for these decisions.
• Providing caveats associated with these decisions.

The Cochrane RevMan template for QES includes guidance for review authors as to what to report under each heading. Of note, PRISMA QES is currently under development. Table 4 presents three currently available tools that have where appropriate been cited in the Cochrane Revman template and guidance and can be used to aid in or guide transparent and clear reporting, eMERGe, ENTREQ and STARLITE. Although the first tool is specifically designed to be used with meta-ethnography some of the principles are applicable to other types of QES. Furthermore, review authors can use generic guidance such as the PRISMA reporting guidelines where applicable to QES (Page, McKenzie et al. 2021, Page, Moher et al. 2021).

Table 4: Tools to aid in transparent and clear reporting of study selection and sampling.

Tool	Relevant items for study selection and sampling
eMERGe guidance: Reporting of meta ethnographies (Suri and Clarke 2009, Finfgeld-Connett and Johnson 2013, France, Cunningham et al. 2019)	#7- Selecting primary studies - describe the screening method (e.g., title, abstract, and/or full text review) and identify who was involved in study selection. - specify the eligibility criteria for study selection (e.g., population, language, year limits, type of publication, study type, methodology, epistemology, country, setting, type of qualitative data, methods, conceptual richness of data, etcetera. - describe any sampling decision for study selection #8- Outcome of study selection - details of numbers of primary studies assessed for eligibility and included. - Reasons for exclusion from comprehensive searches should include numbers of studies screened indicated in a PRISMA figure/flowchart (audit trail) (Page, McKenzie et al. 2021). - Any searching on the basis of theory should describe "reasons for study exclusion and inclusion based on modifications to the review question and/or contribution to theory development" (France, Cunningham et al. 2019). #18- A discussion of the impact of study selection and sampling, the number of included studies / volume of data, upon analysis reflecting on strengths, limitations, and reflexivity.
The Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ statement (Tong, Flemming et al. 2012)	#5- a requirement to clearly describe the information sources used, when the searches were conducted and the rationale for choosing the data sources #6- a description of the literature search #7- a clear and transparent description of the study selection process #9- the review authors "Identify the number of studies screened and provide reasons for study exclusion" - ENTREQ distinguishes between *comprehensive searching* (numbers of studies screened and reasons for exclusion within a figure/flowchart) and *iterative searching* (reasons for study exclusion and inclusion based on modifications of the research question and/or contribution to theory development).
STARLITE (Booth 2006)	Specifically devised for reporting search strategies included the initial letter S for "Sampling strategy"

Figure 4 shows an example of a PRISMA flow diagram template adapted for a QES. This figure is adapted from figure 1 in Page 2021 (Page, Moher et al. 2021).

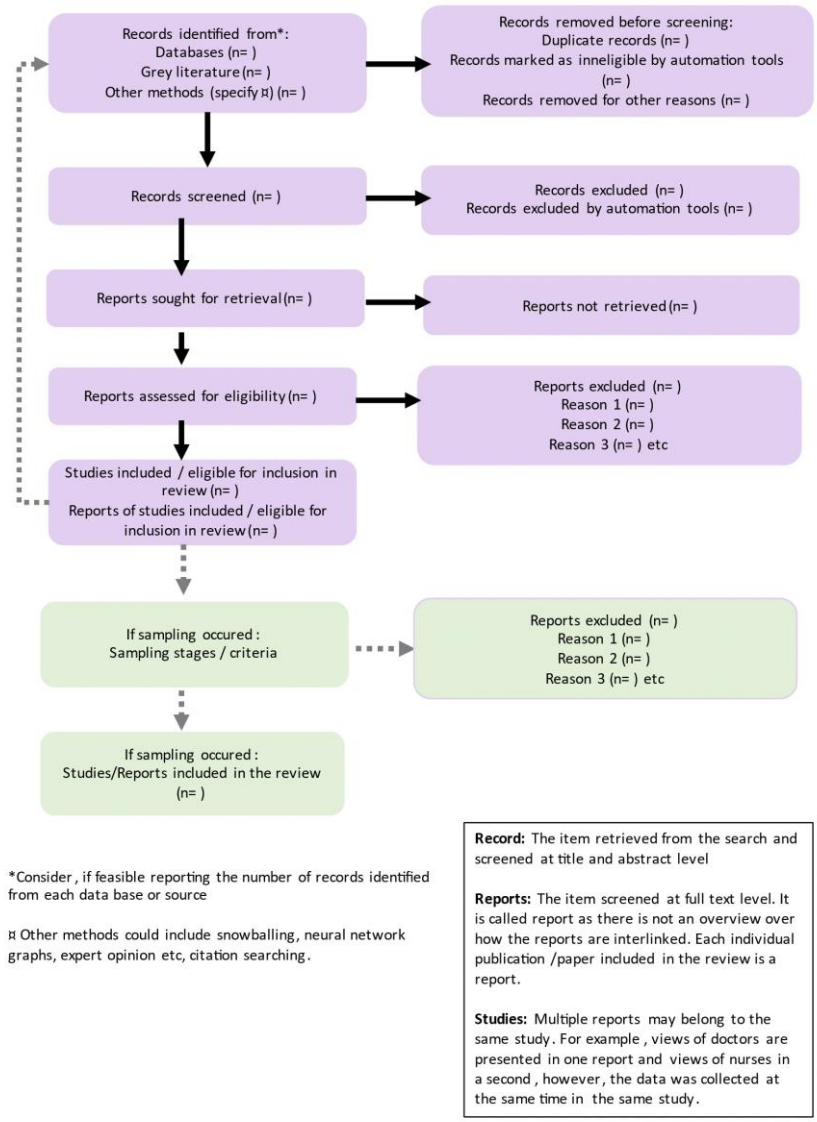


Figure 4: Example of a PRISMA flow diagram template adapted for a QES.

6.7 REFLEXIVITY

Review authors should question how their previous experience with the topic of interest or study population impacted on study selection decisions. For example, might a review author have been more or less likely to sample among studies published within different traditions? How did the review author's individual and collective methodological experience impact the choice of sampling strategy or strategies, and what were the implications of these choices on the synthesised findings?

Collectively, review authors should discuss these issues regularly and create a plan for documentation and reflection.

6.8 Stakeholder engagement and involvement

Stakeholders generally play a vital role in study selection and sampling and their contribution should be reported transparently. Linked with reflexivity considerations, review authors should discuss and think about the involvement of any stakeholders such as commissioners, other researchers, policy makers, patients or public members on study selection and sampling. This is particularly relevant when stakeholders have an interest in seeing particular studies included. If these studies are included, what might the consequences be for the trustworthiness of the synthesis?

Patients and the public are commonly involved in study selection and sampling (Pollock, Campbell et al. 2019, Merner, Lowe et al. 2021). For example, patients and the public can be trained to screen for included studies, collaborate in adjusting the inclusion criteria or making decisions on studies that fall in the grey zone between inclusion and exclusion (Merner 2019, Merner, Lowe et al. 2019). They could also help to choose a sampling method and set up a sampling frame based on their first-hand knowledge of the review topic or their expertise gained by experience.

6.9 Equity, diversity and inclusion

Before beginning study selection and sampling, review authors should consider issues relating to equity, diversity and inclusion. Many commissioners seek to answer questions related to the equity, diversity and inclusion of interventions or how they are implemented. Are equity considerations critical to the planned review? As described in Chapter 1, Equity frameworks such as PROGRESS plus (O'Neill, Tabish et al. 2014) can be used to select and sample studies. PROGRESS-Plus can help review authors to pay explicit attention to established social determinants of health – place of residence, race/ethnicity/culture/language, occupation, gender/sex, religion, education, SES, and social capital in their study selection.

Equity, diversity and inclusion should also be considered when selecting stakeholders (in particular non government organisations patients and the public) to participate in study selection and sampling. Ideally these stakeholders should be selected to represent key equity, diversity and inclusion considerations that are critical for a specific review (for example, women accessing antenatal services in low and middle income countries).

6.10 CHAPTER INFORMATION

Acknowledgements

Thank you to Ashley Muller for critically reading through an early draft and helping to draft the section on automation and machine learning. Thank you to Sara Cooper (Cochrane South Africa, South African Medical Research Council), and Dr Bronwen Merner (La Trobe University) for peer reviewing the chapter, and to the Editors for their input and editorial revisions.

Sources of support

The authors declare no sources of support for writing this chapter.

Conflicts of Interest

Heather Ames was formerly editor with the Cochrane Consumer and Communications group.

Heather Ames, Jane Noyes and Andrew Booth are convenors of the Qualitative and Implementation Methods Group. Noyes is a member of the Cochrane Methods Executive and Editorial Board and contributed to the development of eMERGe. Booth is a convenor of the Information Retrieval Methods Group and contributed to eMERGe and developed STARLIGHT. Ames and Noyes are involved in the development of the data richness/thickness tool.

REFERENCES

Ames, H., E. France, S. Cooper, M. S. Bianchim, S. Lewis, B. Schmidt, I. Uny and J. Noyes (2024). "Assessing qualitative data richness and thickness: development of an evidence-based tool for use in qualitative evidence synthesis Short running title: A data thickness/richness assessment tool."

Ames, H., C. Glenton and S. Lewin (2019). "Purposive sampling in a qualitative evidence synthesis: a worked example from a synthesis on parental perceptions of vaccination communication." BMC Med Res Methodol **19**(1): 26.

Ames, H. M., C. Glenton and S. Lewin (2017). "Parents' and informal caregivers' views and experiences of communication about routine childhood vaccination: a synthesis of qualitative evidence." Cochrane Database of Systematic Reviews(2).

Ames, H. M., C. Glenton, S. Lewin, T. Tamrat, E. Akama and N. Leon (2019). "Clients' perceptions and experiences of targeted digital communication accessible via mobile devices for reproductive, maternal, newborn, child, and adolescent health: a qualitative evidence synthesis." Cochrane Database of Systematic Reviews(10).

Bannach-Brown, A., P. Przybyła, J. Thomas, A. S. Rice, S. Ananiadou, J. Liao and M. R. Macleod (2019). "Machine learning algorithms for systematic review: reducing workload in a preclinical review of animal studies and reducing human screening error." Systematic reviews **8**(1): 1-12.

Benoot, C., K. Hannes and J. Bilsen (2016). "The use of purposeful sampling in a qualitative evidence synthesis: A worked example on sexual adjustment to a cancer trajectory." BMC medical research methodology **16**(1): 1-12.

Bergdahl, E. (2019). "Is meta-synthesis turning rich descriptions into thin reductions? A criticism of meta-aggregation as a form of qualitative synthesis." Nurs Inq **26**(1): e12273.

Booth, A. (2006). "“Brimful of STARLITE”: toward standards for reporting literature searches." Journal of the Medical Library Association **94**(4): 421.

Booth, A. (2016). "Searching for qualitative research for inclusion in systematic reviews: a structured methodological review." Syst Rev **5**: 74.

Booth, A., C. Carroll, I. Ilott, L. L. Low and K. Cooper (2013). "Desperately seeking dissonance: identifying the disconfirming case in qualitative evidence synthesis." Qual Health Res **23**(1): 126-141.

Booth, A., S. Mshelia, C. V. Analo and S. B. Nyakang'o (2019). "Qualitative evidence syntheses: Assessing the relative contributions of multi-context and single-context reviews." Journal of advanced nursing **75**(12): 3812-3822.

Booth, A., J. Noyes, K. Flemming, A. Gerhardus, P. Wahlster, G. J. van der Wilt, K. Mozygemba, P. Refolo, D. Sacchini, M. Tummers and E. Rehfuess (2018). "Structured methodology review identified seven (RETREAT) criteria for selecting qualitative evidence synthesis approaches." J Clin Epidemiol **99**: 41-52.

Brown, S. J., G. J. Carter, G. Halliwell, K. Brown, R. Caswell, E. Howarth, G. Feder and L. O'Doherty (2022). "Survivor, family and professional experiences of psychosocial

interventions for sexual abuse and violence: a qualitative evidence synthesis." Cochrane database of systematic reviews(10).
Callaghan, M. W. and F. Müller-Hansen (2020). "Statistical stopping criteria for automated screening in systematic reviews." Systematic Reviews **9**(1): 1-14.
Campbell, F., M. Holmes, E. Everson-Hock, S. Davis, H. B. Woods, N. Anokye, P. Tappenden and E. Kaltenthaler (2015). "A systematic review and economic evaluation of exercise referral schemes in primary care: a short report." Health technology assessment **19**(60).
Campbell, K., T. Coleman-Haynes, K. Bowker, S. E. Cooper, S. Connelly and T. Coleman (2020). "Factors influencing the uptake and use of nicotine replacement therapy and e-cigarettes in pregnant women who smoke: a qualitative evidence synthesis." Cochrane Database of Systematic Reviews(5).
Campbell, R., P. Pound, M. Morgan, G. Daker-White, N. Britten, R. Pill, L. Yardley, C. Pope and J. Donovan (2011). "Evaluating meta-ethnography: systematic analysis and synthesis of qualitative research." Health Technol Assess **15**(43): 1-164.
Carroll, C., A. Booth, F. Campbell and C. Relton (2020). "What are the implications of Zika Virus for infant feeding? A synthesis of qualitative evidence concerning Congenital Zika Syndrome (CZS) and comparable conditions." PLoS neglected tropical diseases **14**(10): e0008731.
Cochrane Effective Practice and Organisation of Care (EPOC) (2017). Author guidance: EPOC Qualitative Evidence Syntheses guidance on when to sample and how to develop a purposive sampling frame. EPOC Resources for review authors.
Cooper, S., B.-M. Schmidt, E. Z. Sambala, A. Swartz, C. J. Colvin, N. Leon and C. S. Wiysonge (2021). "Factors that influence parents' and informal caregivers' views and practices regarding routine childhood vaccination: a qualitative evidence synthesis." Cochrane Database of Systematic Reviews(10).
DeepL. (2021). "DeepL Translator." Retrieved 17 December, 2021, from https://www.deepl.com/translator.
Dixon-Woods, M., D. Cavers, S. Agarwal, E. Annandale, A. Arthur, J. Harvey, R. Hsu, S. Katbamna, R. Olsen and L. Smith (2006). "Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups." BMC medical research methodology **6**(1): 1-13.
Downe, S., K. W. Finlayson, T. A. Lawrie, S. A. Lewin, C. Glenton, S. Rosenbaum, M. Barreix and Ö. Tunçalp (2019). "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 **17**(1): 76.
Edwards, P., M. Clarke, C. DiGuiseppi, S. Pratap, I. Roberts and R. Wentz (2002). "Identification of randomized controlled trials in systematic reviews: accuracy and reliability of screening records." Stat Med **21**(11): 1635-1640.

Finfgeld-Connett, D. and E. D. Johnson (2013). "Literature search strategies for conducting knowledge-building and theory-generating qualitative systematic reviews." Journal of advanced nursing **69**(1): 194-204.

France, E., I. Uny, R. Turley, K. Thomson, J. Noyes, A. Jordan, L. Forbat, L. Caes and M. S. Bianchim (2023). "A meta-ethnography of how children and young people with chronic non-cancer pain and their families experience and understand their condition, pain services, and treatments." Cochrane Database of Systematic Reviews(10).

France, E. F., M. Cunningham, N. Ring, I. Uny, E. A. Duncan, R. G. Jepson, M. Maxwell, R. J. Roberts, R. L. Turley and A. Booth (2019). "Improving reporting of meta-ethnography: the eMERGe reporting guidance." BMC medical research methodology **19**(1): 1-13.

Gates, A., S. Guitard, J. Pillay, S. A. Elliott, M. P. Dyson, A. S. Newton and L. Hartling (2019). "Performance and usability of machine learning for screening in systematic reviews: a comparative evaluation of three tools." Systematic reviews **8**(1): 1-11.

Gentles, S. J., C. Charles, D. B. Nicholas, J. Ploeg and K. A. McKibbon (2016). "Reviewing the research methods literature: principles and strategies illustrated by a systematic overview of sampling in qualitative research." Systematic Reviews **5**(1): 1-11.

Glenton, C., M. Bohren, S. Downe, E. Paulsen and S. Lewin EPOC qualitative evidence synthesis: protocol and review template, .

Google (2021). "Google Translate."

Higgins, J., J. Thomas, J. Chandler, M. Cumpston, T. Li, M. Page and V. A. Welch (2019). "Chapter 4: searching for and selecting studies." Cochrane Handbook for Systematic Reviews of Interventions: 633.

Lewin, S., C. Glenton and A. D. Oxman (2009). "Use of qualitative methods alongside randomised controlled trials of complex healthcare interventions: methodological study." Bmj **339**.

Lins, S., D. Hayder-Beichel, G. Rücker, E. Motschall, G. Antes, G. Meyer and G. Langer (2014). "Efficacy and experiences of telephone counselling for informal carers of people with dementia." Cochrane database of systematic reviews(9).

Marshall, C. and P. Brereton (2015). Systematic review toolbox: a catalogue of tools to support systematic reviews. Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering.

Merner, B. (2019). Co-producing a Cochrane qualitative evidence synthesis: applying real-world perspectives to full-text screening. Cochrane Colloquium. Santiago, Chile.

Merner, B., D. Lowe, L. Walsh, S. Hill, A. Mussared and C. Wardrope (2019). "Involving stakeholders in Cochrane Review screening." https://community.cochrane.org/news/involving-stakeholders-cochrane-review-screening.

Merner, B., D. Lowe, L. Walsh, A. Synnot, J. Stratil, S. Lewin, C. Glenton, P. von Philipsborn, L. Schonfeld and R. Ryan (2021). "Stakeholder Involvement in Systematic Reviews: Lessons

From Cochrane's Public Health and Health Systems Network." American Journal of Public Health **111**(7): 1210-1215.
Merner, B., L. Schonfeld, A. Virgona, D. Lowe, L. Walsh, C. Wardrope, L. Graham-Wisener, V. Xafis, C. Colombo and N. Refahi (2023). "Consumers' and health providers' views and perceptions of partnering to improve health services design, delivery and evaluation: a co-produced qualitative evidence synthesis." Cochrane Database of Systematic Reviews(3).
Morse, J. M. (2010). "Sampling in grounded theory." The SAGE handbook of grounded theory: 229-244.
Noyes, J., A. Booth and M. Cargo (2019). Chapter 21: Qualitative evidence. Higgins, J.; Thomas, J. Cochrane Handbook for Systematic Reviews of Interventions.
Noyes, J., A. Booth, S. Lewin, B. Carlsen, C. Glenton, C. J. Colvin, R. Garside, M. A. Bohren, A. Rashidian and M. Wainwright (2018). "Applying GRADE-CERQual to qualitative evidence synthesis findings–paper 6: how to assess relevance of the data." Implementation Science **13**(1): 51-61.
Noyes, J., M. Hendry, S. Lewin, C. Glenton, J. Chandler and A. Rashidian (2016). "Qualitative "trial-sibling" studies and "unrelated" qualitative studies contributed to complex intervention reviews." Journal of clinical epidemiology **74**: 133-143.
Noyes, J. and J. Popay (2007). "Directly observed therapy and tuberculosis: how can a systematic review of qualitative research contribute to improving services? A qualitative meta-synthesis." Journal of advanced nursing **57**(3): 227-243.
O'Neill, J., H. Tabish, V. Welch, M. Petticrew, K. Pottie, M. Clarke, T. Evans, J. P. Pardo, E. Waters and H. White (2014). "Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health." Journal of clinical epidemiology **67**(1): 56-64.
O'Mara-Eves, A., J. Thomas, J. McNaught, M. Miwa and S. Ananiadou (2015). "Using text mining for study identification in systematic reviews: a systematic review of current approaches." Systematic reviews **4**(1): 1-22.
Page, M. J., J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl and S. E. Brennan (2021). "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews." Bmj **372**.
Page, M. J., D. Moher, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl and S. E. Brennan (2021). "PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews." bmj **372**.
Patton, M. Q. (2014). Qualitative research & evaluation methods: Integrating theory and practice, Sage publications.
Pawson, R., T. Greenhalgh, G. Harvey and K. Walshe (2005). "Realist review-a new method of systematic review designed for complex policy interventions." Journal of health services research & policy **10**(1_suppl): 21-34.
Pollock, A., P. Campbell, J. Cheyne, J. Cowie, B. Davis, J. McCallum, K. McGill, A. Elders, S. Hagen and D. McClurg (2020). "Interventions to support the resilience and mental health of

frontline health and social care professionals during and after a disease outbreak, epidemic or pandemic: a mixed methods systematic review." Cochrane Database of Systematic Reviews(11).

Pollock, A., P. Campbell, C. Struthers, A. Synnot, J. Nunn, S. Hill, H. Goodare, J. Morris, C. Watts and R. Morley (2019). "Development of the ACTIVE framework to describe stakeholder involvement in systematic reviews." Journal of health services research & policy **24**(4): 245-255.

Popay, J., S. Mallinson, I. Bourgeault, R. Dingwall and R. de Vries (2010). "Qualitative research review and synthesis." The SAGE handbook of qualitative methods in health research: 289-306.

Popay, J., A. Rogers and G. Williams (1998). "Rationale and standards for the systematic review of qualitative literature in health services research." Qualitative health research **8**(3): 341-351.

Porritt, K., J. Gomersall and C. Lockwood (2014). "JBI's systematic reviews: study selection and critical appraisal." AJN The American Journal of Nursing **114**(6): 47-52.

Rohwer, A., L. Hendricks, S. Oliver and P. Garner (2021). "Testing for saturation in qualitative evidence syntheses: An update of HIV adherence in Africa." Plos one **16**(10): e0258352.

Sandelowski, M. (1995). "Sample size in qualitative research." Research in nursing & health **18**(2): 179-183.

Sandelowski, M. and J. Barroso (2006). Handbook for synthesizing qualitative research, springer publishing company.

Shemilt, I., N. Khan, S. Park and J. Thomas (2016). "Use of cost-effectiveness analysis to compare the efficiency of study identification methods in systematic reviews." Syst Rev **5**(1): 140.

Silverman, D. (2013). Doing qualitative research: A practical handbook, Sage.

Suri, H. (2011). "Purposeful sampling in qualitative research synthesis." Qualitative research journal.

Suri, H. and D. Clarke (2009). "Advancements in research synthesis methods: From a methodologically inclusive perspective." Review of Educational Research **79**(1): 395-430.

Thomas, J., A. Noel-Storr, I. Marshall, B. Wallace, S. McDonald, C. Mavergames, P. Glasziou, I. Shemilt, A. Synnot and T. Turner (2017). "Living systematic reviews: 2. Combining human and machine effort." Journal of clinical epidemiology **91**: 31-37.

Thorne, S. (2017). "Metasynthetic madness: what kind of monster have we created?" Qualitative Health Research **27**(1): 3-12.

Tong, A., K. Flemming, E. McInnes, S. Oliver and J. Craig (2012). "Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ." BMC medical research methodology **12**(1): 1-8.

Wainwright, M., C. J. Colvin, A. Swartz and N. Leon (2016). "Self-management of medical abortion: a qualitative evidence synthesis." Reprod Health Matters **24**(47): 155-167.

Wingfield, T., B. Kirubi, K. Viney, D. Boccia and S. Atkins (2023). "Experiences of conditional and unconditional cash transfers intended for improving health outcomes and health service use: a qualitative evidence synthesis." Cochrane Database of Systematic Reviews(3).