STUDY PROTOCOL

Evidence synthesis summary formats for clinical guideline development group members: a mixed-methods systematic review protocol [version 1; peer review: 1 approved with reservations]

Melissa K. Sharp\textsuperscript{1}, Barrie Tyner\textsuperscript{2}, Dayang Anis Binti Awang Baki\textsuperscript{3}, Cormac Farrell\textsuperscript{2}, Declan Devane\textsuperscript{4,5}, Kamal R. Mahtani\textsuperscript{6}, Susan M. Smith\textsuperscript{1}, Michelle O’Neill\textsuperscript{2}, Máirín Ryan\textsuperscript{2,7}, Barbara Clyne\textsuperscript{1,2}

\textsuperscript{1}Health Research Board Centre for Primary Care Research, Department of General Practice, Royal College of Surgeons in Ireland, Dublin, Ireland
\textsuperscript{2}Health Information and Quality Authority, Dublin, Ireland
\textsuperscript{3}School of Medicine, Royal College of Surgeons in Ireland, Dublin, Ireland
\textsuperscript{4}School of Nursing and Midwifery, NUI Galway, Galway, Ireland
\textsuperscript{5}Evidence Synthesis Ireland and Cochrane Ireland, Galway, Ireland
\textsuperscript{6}Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK
\textsuperscript{7}Department of Pharmacology & Therapeutics, Trinity College Dublin, Dublin, Ireland

Abstract

Introduction: Evidence syntheses, often in the form of systematic reviews, are essential for clinical guideline development and informing changes to health policies. However, clinical guideline development groups (CGDG) are multidisciplinary, and participants such as policymakers, healthcare professionals and patient representatives can face obstacles when trying to understand and use evidence synthesis findings. Summary formats to communicate the results of evidence syntheses have become increasingly common, but it is currently unclear which format is most effective for different stakeholders. This mixed-methods systematic review (MMSR) evaluates the effectiveness and acceptability of different evidence synthesis summary formats for CGDG members.

Methods: This protocol follows guidance from the Joanna Briggs Institute on MMSRs and is reported according to the Preferred Reporting Items for Systematic Reviews (PRISMA)-P guideline. A comprehensive search of six databases will be performed with no language restrictions. Primary outcomes are those relating to the effectiveness and preferences for and attitudes towards the different summary formats. We will include qualitative research and randomised controlled trials. Two reviewers will perform title,
abstract, and full-text screening. Independent double-extraction of study characteristics and critical appraisal items will be undertaken using a standardised form. We will use a convergent segregated approach to analyse quantitative and qualitative data separately; results will then be integrated.

**Discussion:** The results of this systematic review will provide an overview of the effectiveness and acceptability of different summary formats for evidence synthesis findings. These findings can be helpful for those in or communicating to guideline development groups. The results can also inform the development and pilot-testing of summary formats for evidence summaries.

**Keywords**
- presentation of findings, evidence summaries, summary of findings table, communication, mixed-methods systematic review

---

**Corresponding author:** Melissa K. Sharp (melissasharp@rcsi.ie)

**Author roles:** **Sharp MK:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Tyner B:** Methodology, Writing – Review & Editing; **Awang Baki DAB:** Writing – Review & Editing; **Farrell C:** Writing – Review & Editing; **Devane D:** Methodology, Writing – Review & Editing; **Mahtani KR:** Methodology, Writing – Review & Editing; **Smith SM:** Methodology, Writing – Review & Editing; **O’Neill M:** Conceptualization, Funding Acquisition, Methodology, Writing – Review & Editing; **Ryan M:** Conceptualization, Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** Health Research Board Ireland [EIA-2019-09].

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2021 Sharp MK et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Sharp MK, Tyner B, Awang Baki DAB et al. Evidence synthesis summary formats for clinical guideline development group members: a mixed-methods systematic review protocol [version 1; peer review: 1 approved with reservations] HRB Open Research 2021, 4:76 https://doi.org/10.12688/hrbopenres.13325.1

**First published:** 15 Jul 2021, 4:76 https://doi.org/10.12688/hrbopenres.13325.1
Introduction

Clinical guidelines support decision making to improve patient outcomes and quality of care in a cost-effective manner. The development of a clinical guideline involves a rigorous synthesis of the best available evidence on a specific clinical topic. It may involve formal consensus methods with a range of multidisciplinary stakeholders. Guideline development groups comprise a range of decision makers, often including healthcare professionals, methodologists, health policymakers, clinicians, and patient representatives— all of whom have varying levels of expertise in evidence synthesis methods. This complicates the consensus process as stakeholders may prioritise and understand the findings of evidence syntheses, such as systematic reviews, differently.

While the methods and recognition of the importance of systematic reviews have advanced in recent decades, there are still barriers to their creation and use. A meta-analysis of nearly 200 systematic reviews registered on the International Prospective Register of Systematic Reviews (PROSPERO) registry found that the average systematic review, from registration to publication date, takes 67.3 weeks, involves an average of five authors, and requires the full-text screening of 63 papers (range: 0–4385). In order for a review to be conducted well, authors must have the time to review the evidence, knowledge to critically appraise study methods, and skills to synthesise the results. Time and resource constraints are becoming an increasing concern as the number of academic papers and systematic reviews being published in recent decades has rapidly increased. This growth accelerated during the recent COVID-19 (coronavirus disease) pandemic, causing an unprecedented “infodemic.” The expansion of evidence, and in decision makers openness to accepting trade-offs in validity, has resulted in the growing popularity of other evidence synthesis methods, such as rapid reviews. This increase in different types of evidence synthesis methods further complicates matters for guideline development groups, who may interpret different types of systematic reviews in different ways based on how familiar they might be with particular approaches.

For those using different types of evidence synthesis to inform clinical guideline development and health policy, the amount of included studies, length, and technical nature of evidence syntheses can make it difficult to find answers about the effectiveness of healthcare interventions. Previous work has highlighted that decision makers more easily understand evidence summaries than complete systematic reviews. These summaries can come in a variety of different formats such as policy briefs, one-page reports, abstracts, summary of findings tables, plain language summaries, visual abstracts or infographics, podcasts, and more. While formatting may vary, decision-makers have expressed several key preferences, such as succinct summaries highlighting contextual factors like local applicability and costs.

Succinctness should be inherent in an evidence summary, but how this distilled information is formatted and presented affects the interpretation and use of systematic reviews. It is currently unclear which evidence summary format is most helpful for decision making for different guideline development group stakeholders. For example, Cochrane recommends a “summary of findings’ table” but testing with users familiar with the Cochrane library and evidence-based practices raised concerns around comprehension and presentation of results and the balance between precision and simplicity. Others have tested the presentation of information using different formats such as an abstract, plain-language summary, podcast or podcast transcription with no clear answer regarding which format was most suited to which stakeholder and resulted in the best understanding. Similarly, infographics, plain-language summaries, and traditional abstracts were found to be equally effective in transmitting knowledge to healthcare providers; however, there were differences in measures of acceptability (i.e., user-friendliness and reading experience).

To better support clinical guideline development groups and decision-makers, it is important to identify which format works best for which stakeholder. Previous reviews have focused on identifying barriers and facilitators to use, or have been solely based on summary of findings tables. As impacts on decision-making and preferences for formats may be evaluated through different study designs, a comprehensive synthesis of the evidence is needed beyond a typical single method systematic review. Mixed methods systematic reviews (MMSR) can more easily identify discrepancies within available evidence, pinpoint how quantitative or qualitative research has focused on particular interest areas, and offer a deeper understanding of findings. A MMSR is especially useful for this project as it brings together findings of effectiveness and experience so findings are more useful for decision makers. Guideline developers need to consider diverse considerations in their work such as feasibility, priority, cost effectiveness, equity, acceptability, and patient values and preferences. Similarly, a MMSR allows us to consider and integrate data from a variety of different questions and synthesize information in a single project.

Objectives

The aim of this mixed methods systematic review is to evaluate the effectiveness of, preferences for, and attitudes towards, different communication formats of evidence summary findings amongst guideline development group members, including healthcare providers, policy makers and patient representatives. To achieve this, the proposed MMSR will answer the following questions:

1. How and to what degree do different summary formats (digital, visual, audio) of presenting evidence synthesis findings impact the end user’s understanding of the review findings?

2. What are the end users’ preferences for and attitudes towards these formats?

Protocol

The proposed systematic review will be conducted in accordance with the Joanna Briggs Institute (JBI) Manual for Evidence...
Synthesis which details the methodology for mixed methods systematic reviews (MMSR).

Eligibility criteria
As this is a MMSR, we will include quantitative (i.e., randomised controlled trials), qualitative, and mixed methods studies evaluating the effectiveness and/or preferences for and attitudes towards evidence summary formats. We will exclude conference abstracts, case reports, case series, editorials, and letters. Further details regarding eligibility criteria are given within the review-relevant sections below.

We are interested in studies involving stakeholders such as policy makers, healthcare providers, and health systems managers, as well as other GDG members such as clinicians, patient representatives, and methodologists such as systematic review authors. We will exclude studies where the sole participants are students, the general population (not involved in the decision-making process), and journalists as communication to these populations is more complex given a wide variety of confounding factors. We will also exclude studies related to clinical decision-making for individual patients.

We have followed the Population, Intervention, Comparison, Outcome (PICO) format for the quantitative review (Table 1) and the Sample, Phenomenon of Interest, Design, Evaluation, Research type (SPiDER) format for the qualitative review (Table 2) and will present unique aspects of each methodological approach within the relevant sections below.

**Quantitative systematic review.** Due to the complexity of stakeholders, evidence synthesis types, and summary formats, there is a high potential that confounding factors will be extensive. Relatedly, randomised controlled trials (RCTs) are the most appropriate design to evaluate the effectiveness of the interventions in question. Thus, we chose to restrict to RCTs only in order to focus on the performance and impact of summary formats in optimal settings. We will include studies where the intervention is any summary mode (e.g., visual, audio, text-based) which communicates the findings from an evidence synthesis study (e.g., systematic review, qualitative evidence synthesis, rapid review) to policy-makers and decision makers, including guideline development groups (GDGs). We anticipate that included summary formats will encompass visual abstracts, Summary of Findings tables, one-page summaries, podcasts, Graphical Overview of Evidence Reviews (GoFER) diagrams, and others. Studies in which the summaries are one component of a multi-component intervention will be excluded, as will decision aids for direct patient care.

For studies examining the effectiveness of evidence summary formats, we will include any comparison to an alternative active comparator. Studies where the comparison is no intervention (e.g., the plain full-text of a manuscript) will be excluded, though we do not anticipate finding evidence syntheses with no form of summary or abstract.

Our primary outcomes of interest are:

1. **Effectiveness**
   a. User understanding and knowledge, and/or beliefs in key findings of evidence synthesis (e.g., changes in knowledge scores about the topic included in the summary)

<table>
<thead>
<tr>
<th>Table 1. PICO for the quantitative review of effectiveness.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population</strong></td>
</tr>
<tr>
<td><strong>Intervention</strong></td>
</tr>
<tr>
<td><strong>Comparator</strong></td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2. SPiDER for the qualitative evidence synthesis.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sample</strong></td>
</tr>
<tr>
<td><strong>Phenomenon of Interest</strong></td>
</tr>
<tr>
<td><strong>Design</strong></td>
</tr>
<tr>
<td><strong>Evaluation outcomes</strong></td>
</tr>
<tr>
<td><strong>Research type</strong></td>
</tr>
</tbody>
</table>
b. Self-reported impact on decision-making

c. Intervention metrics (e.g., the time needed to read the summary, expressed language accessibility issues or scale scores)

2. Acceptability

a. Preferences and attitudes (e.g. Likert scales reporting user satisfaction, perceptions, readability).

We will not be including outcomes related to health literacy, numeracy, nor risk communication in patient-centred care.

**Qualitative evidence synthesis.** Primary studies investigating the understanding and acceptability of evidence summary formats will include qualitative studies (e.g. interviews or focus groups). Mixed-methods studies with primary qualitative data collection will be included if they meet the inclusion criteria for a randomised controlled trial and where it is possible to extract the findings derived from the qualitative research.

Our primary outcomes of interest relate to participant’s views and experiences with summary formats. This includes their perceptions of the impact of summary formats on their understanding, knowledge, and decision making, and participant’s beliefs, attitudes, and feelings towards usability and readability.

**Information sources and search strategy**

The following databases will be searched from inception to May 2021: Ovid MEDLINE, EMBASE, APA PsycINFO, CINAHL (Cumulative Index to Nursing and Allied Health Literature), Web of Science, and Cochrane Library. The search strategy for Ovid MEDLINE includes a combination of keywords and medical subject headings (MeSH) terms for GDG members, evidence syntheses, and formats for the communication of findings (see Table 3). As we are looking for primary research on the impacts or effects of interventions and

<table>
<thead>
<tr>
<th>Table 3. Ovid MEDLINE search strategy.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ovid MEDLINE(R) and Epub Ahead of Print, In-Process, In-Data-Review &amp; Other Non-Indexed Citations, Daily and Versions(R) 1946 to April 13, 2021</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>15</td>
</tr>
<tr>
<td>16</td>
</tr>
<tr>
<td>17</td>
</tr>
<tr>
<td>18</td>
</tr>
</tbody>
</table>
attitudes towards them, we do not anticipate that this literature will be found in grey literature sources such as government or agency websites. Additionally, it is anticipated that controlled trials will have short time points of assessment (and follow-up) thus we do not believe that searching registries will benefit our study. This search strategy has been informed by the strategies of similar reviews in the same topic area.\textsuperscript{16,23} Aligned with the Peer Review of Electronic Search Struggles (PRESS) Statement\textsuperscript{28}, we engaged a medical librarian after the MEDLINE search was drafted but before it was translated to the other databases. As we are including a range of study designs, we did not apply study design specific filters. Although we have used a PICO and SPiDER approach for the quantitative and qualitative reviews we used the PICO format to inform the search strategy as previous researchers found that the SPiDER approach for search strategies may be too restrictive and specific\textsuperscript{31}. Language and date restrictions will not be applied.

Backwards citation identification on all eligible studies will be performed using the citationchaser Shiny application built in R version 1.4\textsuperscript{32}. This application performs backwards citation screening (reviewing reference lists) and internally de-duplicates results. Each step of the search is summarised for transparency and references are given as a downloadable RIS file.

Data management and selection process
All citations will be downloaded and stored in Zotero reference manager version 5.0. For ease, rather than using Zotero for screening, title and abstract screening will be managed using Covidence. Two reviewers will independently screen titles and abstracts for inclusion criteria. Disagreements for inclusion will be resolved through discussion. If it is still unclear if the paper should be included, both authors will review the full version of the paper and discuss it again. If there is still disagreement, a third review author will be consulted.

The screening process will be documented in the final manuscript using the Preferred Reporting Items for Systematic Reviews (PRISMA) flow diagram\textsuperscript{33} and a supplemental file detailing the reason for exclusion for each individual study will be made publicly available.

Data collection
Two review authors will independently extract data from each of the included studies using a standardised data-extraction form. If there are disagreements or discrepancies, the two authors will discuss and consult with a third review author if needed. Where possible, qualitative outcomes such as themes and categories will be extracted into the standardized form, however, included articles containing qualitative methods will be also imported in NVivo12 for line-by-line coding for information related to outcomes. This separate but parallel data extraction is important for our analytical approach of the qualitative data which is discussed in greater detail in the Qualitative Analysis section. The following data will be extracted:

- Bibliometric data (first author, title, journal, year of publication, language)
- Study characteristics (setting, participants demographics, country, study design, intervention, comparators, theoretical framework, analytical approach)
- Intervention characteristics will be collected following the structure of the Template for Intervention Description and Replication (TiDieR) checklist\textsuperscript{34} to provide detailed information on the why, what, who, how, where, and when of the intervention described.
- Outcomes (quantitative estimates of effectiveness and acceptability; qualitative expressions of views, attitudes, opinions, experiences, perceptions, beliefs, feelings, and understanding)
- Data to critically appraise included studies (JBI critical appraisal tools)
- Funding sources

If information is missing from the study report, we will contact authors to inquire about these gaps. We will provide narrative synthesises in lieu of imputing missing data.

Bias and quality assessments
The JBI critical appraisal checklists will be used for randomized controlled trials and qualitative research. As data from separate quantitative and qualitative evidence is integrated in a MMSR, the JBI does not recommend the GRADE nor ConQual approaches for assessing overall confidence in synthesized findings\textsuperscript{26}. Two review authors will independently complete the critical appraisal checklist for each included study. Differences will be resolved through discussion and consultation with a third review author if necessary.
If quantitative data allows for a meta-analysis, a forest plot will be generated using R. If we find a low number of studies, large treatment effects, few events per trial, or all trials are of similar sizes, we will use the Harboard test for publication bias\(^{39}\) as it reduces the false positive rate. Egger’s test\(^{40}\) for funnel plot asymmetry will be used to investigate small study effects and publication bias.

### Quantitative analysis

A narrative synthesis will be performed, however, if appropriate, quantitative data from randomised control trials will be synthesised using meta-analysis. Heterogeneity will first be explored by assessing the study characteristics that may vary across the included studies (for example, participant group, study design, risk of bias, interventions or outcomes). If sufficient data is available, subgroup analyses (e.g. participant groups such as medical professionals versus policy makers or intervention type such as visual abstracts versus plain abstracts) will be conducted. Furthermore, statistical heterogeneity will be explored according to statistical guidance on heterogeneity\(^3\), an estimated \(I^2\) of 50-90% represents substantial heterogeneity. We will weigh this against an \(\chi^2\) test for heterogeneity (<.10). If our results indicate 50% or greater and a low \(\chi^2\) statistic, this indicates that the heterogeneity may not be due to chance, thus we will not pool results into a meta-analysis. If data can be pooled, effect sizes and accompanying 95% confidence intervals will be reported as either relative risks (for dichotomous and dichotomised ordinal data) or standardized mean differences (for continuous data).

### Qualitative analysis

Where possible, qualitative findings will be pooled together using the meta-aggregation approach, which allows a reviewer to present findings of included studies as originally intended by the original authors\(^7\). This approach organises and categorises findings based on similarity in meaning and avoids re-interpretation. Therefore, it does not violate paradigms and approaches used by the original study authors. This approach also enables meaningful generalizable recommendations for practitioners and policy makers\(^{18}\). If textual pooling is not available, a narrative summary will be presented.

### Mixed methods synthesis

Following JBI guidance for MMSR, we will use a convergent segregated approach that conducts separate quantitative and qualitative syntheses and then integrates the findings of each\(^{36,39}\). This integration process is called ‘configuration’ and allows one to arrange complementary evidence into a single line of reasoning. After separate analyses are conducted, they will be organized into a coherent whole, as they cannot be directly combined nor can one refute the other\(^{40}\). Data will be triangulated during the interpretation stage to identify areas where there is convergence, inconsistency, or contradiction in the data. If we have a sufficient number of included studies, subgroup analyses will be performed to investigate differences based on participant groups (e.g., clinicians versus policy makers) and outcomes (e.g., understanding, acceptability, etc.). Within this approach, we will use thematic synthesis. This method codes text ‘line-by-line’, develops descriptive themes, and finally consolidates and generates analytical themes\(^{41,42}\). Initial coding will be performed independently by two authors who will meet and discuss similarities and differences in coding to start grouping them into a hierarchical tree structure of descriptive themes. A drafted summary of findings will be created by one author, reviewed by both, and discussions will be held until a final version is agreed upon. Two authors will discuss the descriptive themes and, as a group, will draft the final analytical themes with accompanying detailed descriptions.

### Registration and amendments

As the focus of this review is not evaluating health-related interventions nor outcomes, we will not register the protocol on PROSPERO. However, we will preregister the study on Open Science Framework. If an amendment to this protocol is necessary, the date of each amendment will be given alongside the rationale and description of the change(s). This information will be detailed in an appendix accompanying the final systematic review publication. Changes will not be incorporated into the protocol.

### Dissemination of information

Findings will be disseminated as peer-reviewed publications. Data generated from the work proposed within this protocol will be made available on the aforementioned OSF project page.

### Discussion

This review will summarise the evidence on the effectiveness and acceptability of different evidence synthesis summary formats. By including a variety of evidence summary types and stakeholder participants, results can help tease apart the real-world complexity of guideline development groups and provide an overview of what summary formats work for which stakeholders in what circumstances. It is expected that review findings can support decision-making by policy-makers and GDGs, by establishing the best summary formats for presenting evidence synthesis findings.

### Data availability

No data are associated with this article.

### Reporting guidelines

OSF: PRISMA-P checklist for ‘Evidence synthesis summary formats for clinical guideline development group members: a mixed-methods systematic review protocol’. [https://doi.org/10.17605/OSF.IO/SK4NX]\(^{43}\)

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

### Acknowledgements

We would like to thank Paul Murphy for his help with developing the search strategy and translating it to other databases.


Open Peer Review

Current Peer Review Status: ?

Karin Hannes ¹
Research group SoMeTHin'K (Social, Methodological and Theoretical Innovation Kreative), CESO, Faculty of Social Sciences, KU Leuven, Leuven, Belgium

This is a review underpinned by clear and interesting questions, that warrant a mixed method strategy. There is room for improvement though on the level of motivating choices for concepts and fine-tuning methods.

The biggest issue I have is on the relation between content and form in the choice for the mediums to be compared. The authors tend to compare forms (?), but attitude towards form is not fully independent to the way messages are framed, nor by whom they are disseminated. How do the authors intend to deal with this in their comparison? (e.g. people might prefer a form such as video but it might fail to bring the message across, as participants are distracted by the form).

I was not fully sure why things like surveys measuring attitudes and opinions of people were not taken along in this review. (I am aware of the fact that guidance on the inclusion of such designs in reviews is scarce though-consider it a free offer to work towards a more comprehensive review type). Particularly, because in the RCTs authors tend to rely on self-reported measures, with LIKERT scales. Here, a comparison could sit with time, rather than comparator. Does people’s attitude change when exposed to different formats over time?

Sometimes the arguments are not fully clear. In the introduction, attention is paid to problems complicating the conduct of systematic reviews. This is not the core focus of the project. The project is linked to the translation phase of turning review findings into evidence summaries. This is a matter of content, form and channels rather than the SR pre-track issues. It would therefore be more appropriate to highlight the diversity of actors involved in producing guidelines in relation to how communication and dissemination channels have changed because of that. I am a big fan of the authors discussing things like podcasts and multimedia accounts as examples. Naturally, this is linked to a society moving into multimodality and the limitations of numbers and narratives in reaching out to or working with different publics.
The PICO would perhaps work better if clinicians is changed into professionals. It would make it relevant to a broader target group. Also, the comparator could also be no summary at all, for example, reading findings from reviews straight into the guideline procedure. Why are these explicitly excluded from the review?

Do the authors only include qualitative studies where an intervention is evaluated by participants in terms of meaningfulness or applicability? Or do they also include qualitative studies where attitudes, viewpoints and options for form/channel/content are discussed from a more theoretical point of view, or in pilot cases? It is not clear where the selection sits on the qualitative side, but the data collection part suggests the first option.

Perhaps the authors may want to bring a rationale for why they want to limit the formats to textual and audio-visual productions, as there are many more forms available through which evidence is and can be communicated (such as installations, theatre, etc.) Perhaps these forms haven't been studied extensively in the context of guidelines, but in a protocol phase we are not sure what to find yet :-).

I would find it more practical if the authors would discuss their position towards critical appraisal and how to structure this instead of putting it in the data collection part as an extraction category (it is not clear what will actually be extracted for what purpose). Also, critical appraisal checklist for primary studies have a different purpose than GRADE and CONQUAL that work on a review level (which is clearly not the focus here when talking about the pool of studies included). I think the authors need to spell the two out in more detail, as they are used in different phases of the review process.

Would quality of a study not rather fit a sensitivity analysis rather than a subgroup analysis? What purpose would it have as a subgroup in the review? It has disappeared in the mixed method part as a subgroup analysis, so wouldn't mention it as a subgroup in other parts of the text.

What do the authors mean with the phrase "If textual pooling is not available, a narrative summary will be presented." Is a narrative summary not already a textual pooling? Unclear.

While configuration might have different meanings, it is generally not commonly used as a metaphor for a separate type of synthesis. Configuration would make the different parts unrecognizable in the whole. I don't think meta-aggregation would achieve this level of integration. Perhaps integration is better than configuration here.

Also, thematic synthesis is not the same as meta-aggregation, and line by line coding is certainly not a part of meta-aggregation. I believe the authors need to make an informed choice between qualitative evidence synthesis approaches and then follow the guidance through.

Is the rationale for, and objectives of, the study clearly described?
Yes

Is the study design appropriate for the research question?
Yes
Are sufficient details of the methods provided to allow replication by others?  
No

Are the datasets clearly presented in a useable and accessible format?  
Not applicable

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Methods

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.