The effectiveness and acceptability of evidence synthesis summary formats for clinical guideline development groups: A mixed-methods systematic review

Melissa Kathleen Sharp (melissasharp@rcsi.com)
Royal College of Surgeons in Ireland https://orcid.org/0000-0001-5261-1573
Dayang Anis Binti Awang Baki
Royal College of Surgeons in Ireland School of Medicine
Joan Quigley
Health Information and Quality Authority
Barrie Tyner
Health Information and Quality Authority
Declan Devane
NUI Galway School of Nursing and Midwifery
Kamal R. Mahtani
University of Oxford Nuffield Department of Primary Care Health Sciences
Susan M. Smith
Trinity College Dublin Faculty of Health Sciences
Michelle O'Neill
Health Information and Quality Authority
Mairin Ryan
Health Information and Quality Authority
Barbara Clyne
: Royal College of Surgeons in Ireland Department of General Practice

Research Article

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

Posted Date: May 12th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1646845/v1

Version of Record: A version of this preprint was published at Implementation Science on October 27th, 2022. See the published version at https://doi.org/10.1186/s13012-022-01243-2.

Abstract

Introduction

: Clinical guideline development often involves a rigorous synthesis of evidence involving multidisciplinary stakeholders with different priorities and knowledge of evidence synthesis; this makes communicating findings complex. Summary formats are typically used to communicate the results of evidence syntheses, however, there is little consensus on which formats are most effective and acceptable for different stakeholders.

Methods

This mixed-methods systematic review (MMSR) aimed to evaluate the effectiveness and acceptability (e.g., usability preferences, and attitudes and preferences towards) of evidence synthesis summary formats for GDG members. We followed the PRISMA 2020 guideline and Joanna Briggs Institute Manual for Evidence Synthesis for MMSRs. We searched six databases (inception to April 20, 2021) for randomised controlled trials (RCTs), RCTs with a qualitative component, and qualitative studies. Screening, data extraction, and quality appraisal was performed in duplicate. Qualitative findings were synthesised using meta-aggregation and quantitative findings are described narratively.

Results

We identified 17,240 citations and screened 54 full-text articles, resulting in 22 eligible articles (20 unique studies): 4 articles reported the results of 5 RCTs, one of which also had a qualitative component. The other 18 articles discussed the results of 16 qualitative studies. Therefore we had 5 trials and 17 qualitative studies to extract data from. Studies were geographically heterogeneous and included a variety of stakeholders and summary formats. All 5 RCTs assessed knowledge or understanding with 3 reporting improvement with newer formats. The qualitative analysis identified 6 themes: ‘presenting information’, ‘tailoring content for end users’, ‘trust in synthesis producers’, ‘knowledge requirements’, ‘quality of included studies’, and ‘properly contextualising findings’. Across these themes, the synthesis resulted in 130 recommendations for practice. Nine recommendations were supported by both quantitative and qualitative evidence and 121 by only qualitative. A majority focused on how to present information (n = 68) and tailor content for different end users (n = 24).

Conclusions

This MMSR provides guidance on how to improve evidence summary structure and layout. This can be used by synthesis producers to better communicate to GDGs. Study findings will inform the co-creation of evidence summary format prototypes based on GDG member’s needs.

Word Count:

345

Registration:

The protocol for this project was previously published and the project was pre-registered on Open Science Framework. [1, 2]

Contributions To The Literature

- Summaries are often used to communicate evidence synthesis findings, however, there is no consensus on the most effective way to communicate or what works for different audiences.
- This review explored the effectiveness and acceptability of different summary formats for different audiences.
- We identified recommendations to help evidence synthesis producers better communicate to different audiences. These include guidance on formatting, tailoring content for end users, instilling trust in the work, establishing and helping knowledge requirements,

detailing the quality of included studies, and properly contextualising findings.
- Results will guide the creation of summary formats better tailored to end user's needs).

Background

Clinical guidelines are an important tool for the practice of evidence-based medicine. Often involving rigorous syntheses of the best available evidence, clinical guidelines (CG) aim to improve healthcare in a cost-effective manner by assisting decision-making for clinicians and policy makers. [3–5] Guideline development groups (GDG) are comprised of a multidisciplinary decision makers such as healthcare professionals, methodologists, and patient representatives. These participants engage in the guideline development process which may involve formal consensus methods amongst these stakeholders – all with different priorities and understanding of evidence synthesis methods. [6]

There are many barriers to clinical guideline development and implementation including a lack of time, money, expertise, awareness, or clarity on local applicability. [7–9] Guideline developers need to consider a variety of factors throughout the process such as feasibility, cost effectiveness, equity, acceptability, and patient preferences, alongside the research evidence. [10,11] Evidence syntheses, such as systematic reviews, may be infrequently used by health care managers and policy makers due to intrinsic factors such as format and content and extrinsic factors such as lack of awareness and skills to seek, appraise, and interpret systematic reviews. [12,13] Review or evidence synthesis summaries have been proposed as a way to improve the uptake and usefulness of evidence syntheses for decision makers. [12,13]

Evidence synthesis summaries can come in a variety of different formats such as one-page plain language reports, policy briefs, summary of findings tables, visual abstracts or infographics, and more. While summaries may be more easily understandable than complete systematic reviews [14,15], review summaries and overviews are often too long and complex, and may require additional work to effectively 'translate' the evidence for policymakers, [16] There are a wide variety of summaries available with no clear consensus on the most effective format. [14,15]

It is critical to identify the best summary formats to ensure effective communication with multidisciplinary GDGs as they interpret evidence syntheses and develop clinical guidelines to support evidence-based decision making. [17] This study aimed to evaluate the effectiveness of, and acceptability of (e.g., preferences for, and attitudes towards) different communication formats of evidence synthesis summary formats amongst GDG members. To support a multifaceted view on the guideline development process, we conducted a mixed methods systematic review (MMSR) as this method offers a deeper understanding of findings, more easily identifies discrepancies in the evidence and is more useful for decision makers. [18,19] An MMSR approach was chosen as we wanted to examine different aspects of a particular phenomenon – i.e., the effects that summary formats may have on knowledge or decision-making and how acceptable these formats were to users. [20]

Methods

We conducted a mixed methods systematic review (MMSR) according to a pre-registered and published protocol. [1,21] The MMSR followed the guidance of the Joanna Briggs Institute (JBI) Manual for Evidence Synthesis, using a convergent segregated approach, [19] and the PRISMA 2020 checklist (Appendix 1) [22].

Study designs and eligibility criteria

Eligible studies were included if they were randomised controlled trials (RCTs) comparing alternative summary formats for evidence syntheses, randomised controlled trials with a supplemental qualitative component, or qualitative studies such as focus groups, interviews, or open-ended surveys. Eligible participants were those who could be involved in clinical guideline development groups (e.g., healthcare professionals, policy makers, patient representatives, etc.) and outcomes related to effectiveness, acceptability (e.g., views and preferences) of summary formats. We did not include observational studies as there is a high potential that confounding factors will be extensive due to the complexity of stakeholders, evidence synthesis types, and summary formats involved. We also did not include studies involving students, journalists, or the general public as communication to these populations is more complex. Further details can be found in the protocol. [21]

Search strategy and study selection

We searched six databases, Ovid MEDLINE, Embase MEDLINE (Medical Literature Analysis and Retrieval System Online), APA (American Psychological Association) PsycINFO, CINAHL (Cumulative Index to Nursing and Allied Health Literature), Web of Science, and Cochrane Library, from inception to April 20, 2021 (Appendix 2). All titles, abstracts and full-texts were independently double screened (DAB, BC, JQ, MKS, BT) using Covidence. [23] Disagreements were discussed between two lead reviewers (BC, MKS) until consensus was achieved. The complete list of eligible articles and potentially relevant studies with exclusion justifications are available on the project's OSF page. [1]. We used the CitationChaser Shiny application to perform backwards citation identification. [24,25] One reviewer (MKS) manually screened citations that the app was unable to include (e.g. reports without a DOI).

Data extraction and appraisal of studies

The data extraction form was piloted by two reviewers (MKS, DAB) on one article, required changes were discussed, and the final data extraction was performed using this form and the TiDiER checklist. [26] Study quality was assessed using the JBI Critical Appraisal Checklist for Qualitative Research and the JBI Checklist for RCTs as appropriate [27]. All data extraction was performed independently in duplicate (DAB, BC, JQ, MKS). Disagreements were discussed with the lead author (MKS) and resolved by consensus. Data extraction templates are available on OSF. [1]

Analysis and synthesis of findings

As we did not have a sufficient number of quantitative studies included, we were unable to perform some analyses as described in the protocol (e.g., a meta-analysis, the Harboard test for publication bias [28], Egger's test [29], and statistical heterogeneity [30]). As established in our protocol, since we could not perform a meta-analysis, a narrative synthesis was performed.

Qualitative findings were synthesised using the pragmatic meta-aggregation approach which allows a reviewer to present findings of included studies as originally intended by the original authors. [31,32] Meta-aggregation seeks to enable generalisable statements in the form of recommendations to guide practitioners and policy makers. Findings (defined as a verbatim extract of the author's analytical interpretation of the results or data) from the results section of manuscripts and accompanying illustrations (direct quotations or statements from participants) were coded as 'unequivocal evidence.' Findings with no illustrations or an illustration lacking clear association were 'equivocal/credible.' Findings which were not supported by the data were 'unsupported'. [31,33] NVivo 12 was used to analyse results from primary qualitative studies and accompanying illustrations. [34] One author (MKS) performed the initial line-by-line coding of equivocal, unequivocal, and unsupported findings which was checked by a second reviewer (BC). [19,35] These findings were then synthesized into categories, based on similarity in meaning, then distilled further into actionable recommendations for practice following the same coding and reviewer process (MKS, BC). As recommended by JBI, we did not differentiate between equivocal and unequivocal findings when aggregating them into categories. These coding steps with results are detailed in Figure 1.

To synthesise findings from both qualitative and quantitative evidence, we followed the JBI guidance for MMSR and used a convergent segregated approach as we conducted separate quantitative and qualitative syntheses and then integrated the findings of each. [19,36] We juxtaposed the synthesised quantitative and qualitative findings then organized the linked the findings in a single line of reasoning to produce an overall configured analysis. [20] This integration process identifies areas of convergence, inconsistency, or contradiction. [37] (Appendix 4). The final table of recommendations was agreed upon by the entire multidisciplinary author team.

Results

Search results

After deduplication of identified records, we screened 17,240 titles and abstracts and 54 full-text articles. We identified 22 articles for inclusion which all underwent backwards citation screening. The screening process is shown in the PRISMA Flow Diagram (Figure 2). The search strategy output and reasons for inclusion/exclusion files are available on OSF. [1] Of note, many studies had multiple phases or participant groups. We included the study if we could clearly separate the methods and results for the phase and/or group. Where possible we extracted information only from the eligible phase/group.

Characteristics of included studies

Our final sample included 22 full-text articles representing 20 unique studies. This included 16 qualitative studies, 4 RCTs, and 1 mixed-methods RCT and qualitative study (Tables 1 and 2) involving 908 total participants from a variety of different stakeholder groups (Table 1). Many studies involved a multidisciplinary mix of participants such as researchers, health professionals, and policymakers [38–52] although some had homogenous groups of clinicians [53–55] or decision-makers. [56–58] The majority of types of evidence syntheses were systematic reviews but one study related specifically to network meta analyses (NMA), one to diagnostic test accuracy (DTA) reviews and one to updating reviews. Seven studies involved an international mix of participants [44,47,51,52,56,58,59]; five were from Canada [40,43,54,55,57], three from the United States of America [41,42,45,48,49], two from Croatia [39,50], two from England [38,53], and one from Kenya [46]. Most were funded by national agencies [39,40,42,42,43,45,48–50,54–57] such as the Canadian Institutes of Health Research [40,43,54,55] or the Agency for Healthcare Research and Quality [42,45,48,49,57].

The TiDiER checklist was used to gather intervention data detailed in Tables 1, 2 and 3. The majority of included qualitative studies conducted either focus groups [39,40,45,54,55] or one-on-one semi-structured interviews. [38,41–43,46–49,53,56–60] (Table 1) RCTs were conducted either with an online survey [50,51] or through in person workshops (Tables 2 and 3). [46,52] There were a wide variety of summary formats tested including de novo summary prototypes [40,43,45,46,53–55,57,58], Grading of Recommendations, Assessment, Development and Evaluations (GRADE) Summary of Findings (SoF) evidence tables [44,46,47,56,59], MAGICapp [48,49], Tableau. [48,49], evidence flowers [38], plain language summaries [39], and infographics [39]. Summary formats covered a wide variety of clinical topics. (Tables 1 and 2)

Quality appraisal

We found the quality of reporting for the qualitative studies was quite poor. The main weakness across these studies included not providing information on philosophical perspectives (11/17) [38–43,45,46,48,49,55,57,58], not locating the researcher culturally or theoretically (15/17) [38,39,43–47,49,53–59] and not addressing the influence of the researcher on the research (15/17) [38,39,41–49,54–59]. Several interviews or focus groups also did not provide clear direct quotes from participants (6/17) [40,43,45,48,49,55,60]. On the other hand, the four quantitative studies were mostly reported clearly with low risk of bias. [46,50–52] The main weaknesses related to descriptions of the blinding of treatment assignment for the outcome assessors and those delivering treatment (2/4) [46,52]. Completed JBI critical appraisal checklists can be found in Appendix 3.

Quantitative analysis

The summary formats tested across the five included RCTs (described across four papers) are described in detail in Table 3. Four RCTs compared alternative versions of SoF tables against a format in current practice and/or a standard systematic review. [46,51,52] One study compared an infographic to a plain language summary (PLS) and scientific abstract (SA). [50] Studies were largely multidisciplinary and results were not presented by stakeholder group. An exception to this was the study by Buljan et al. 2018 which conducted separate trials with patient representatives ('consumers') and doctors. There were no differences between the groups in knowledge scores for both the plain-language summary (PLS) and infographic formats. However, patient representatives reported lower satisfaction (user-friendliness) and reading experience with both formats when compared to doctors. As the quantitative studies used a variety of scales and summary formats, we could only summarise results narratively.

In preparation for the mixed methods synthesis, we identified 74 individual findings from quantitative studies (Appendix 4) and synthesised these into four main areas which related to review outcomes of Knowledge/Understanding, Satisfaction/Reading Experience, Accessibility/Ease of Use, and Preference. (Figure 1). These individual findings helped identify areas of convergence, inconsistency, or contradiction with the qualitative findings and recommendations described later.

<u>Knowledge or Understanding</u>

All five RCTs assessed knowledge or understanding as an outcome (Table 4). No studies employed standardised measures, choosing to use study specific questions. Two articles, reporting the results of three studies, found that the new format improved knowledge or understanding. [51,52] Carasco-Labra et al. reported that compared to a standard SoFs table, a new format of SoF table with seven

alternative items improved understanding. [51] Of seven items testing understanding, three showed similar results, two showed small differences favoring the new format, and two (understanding risk difference and quality of the evidence associated with a treatment effect) showed large differences favoring the new format [63% (95% CI: 55, 71) and 62% (95% CI: 52, 71) more correct answers, respectively]. In two small RCTs, Rosenbaum et al. found that the inclusion of a SoF table in a review improved understanding and rapid retrieval of key findings compared to reviews with no SoF table. [52] In the second RCT, there were large differences in the proportion that correctly answered questions about risk in the control group (44% vs. 93%, P=0.003) and risk in the intervention group (11% vs. 87%, P<0.001). Two studies reported no significant differences between formats in knowledge or understanding. [46,50]

Ease of use/Accessibility

All five RCTs provided some assessment of ease of use and accessibility, measured in a variety of ways (Table 4). Buljan et al. reported that user friendliness was higher for an infographic compared to a PLS for doctors and patient representatives [Patients median infographic score: 30.0 (95% CI: 25.5–34.5) vs. PLS: 21.0 (19.0–25.0); Doctors median infographic score: 36.0 (30.9–40.0) vs. PLS: 29.0 (26.8–36.2)]. [50] while Carasco-Labra et al. reported that in six out of seven domains, participants rated information in the alternative SoF table as more accessible overall (MD 0.3, SE 0.11, P=0.001). [51] Opyio et al's graded-entry SoF formats were associated with a higher mean composite score for clarity and accessibility of information about the quality of evidence (adjusted mean difference 0.52, 95% CI 0.06 to 0.99). [46] In two small RCTs, Rosenbaum et al. found that participants with the SoF format were more likely to respond that the main findings were accessible. [52] The second RCT demonstrated, that in general, participants with the SoF format spent less time finding answers to key questions than those without.

Satisfaction

Two studies assessed satisfaction (Table 4). Buljan et al. reported that both patients and doctors rated an infographic better for reading experience than a PLS, even though it didn't improve knowledge [Patients median infographic score: 33.0 (95% CI: 28.0 – 36.0) vs. PLS: 22.5 (19.0 – 27.4); Doctors median infographic score: 37.0 (26.8 – 41.3) vs. PLS: 24.o(21.3 – 27.2)] [50] Carasco-Labra et al. reported that participants were more satisfied with the new format of SoF tables. (5/6 questions where the largest proportion was in favour of alternate SoF tables). [51]

Preference

Two studies assessed user preference (Table 4). Carasco-Labra et al. reported that participants consistently preferred the new format of SoF tables (MD 2.8, SD 1.6). [51] Similarly, Rosenbaum et al. reported that overall participants preferred the alternative (or new) format of SoF tables compared to the current formats (MD/SD: 2.8/1.6). [52]

Qualitative analysis

From 16 qualitative studies and 1 RCT with a supplemental qualitative component, line by line coding identified 542 equivocal and unequivocal findings within the results section of the articles. No unsupported findings were identified. (Figure 1) From these initial 542 findings, we synthesized them further into 393 findings across 6 categories defined as follows:

1. Presenting information (comments on the content, structure, and style of the summary format);
2. Tailoring information for end users (inherently linked to the presentation of information but more focused on accommodating end user's different learning styles, backgrounds, and needs and appropriately tailoring content);
3. Contextualising findings (properly framing the findings themselves within the relevant context by providing information such as setting, cost constraints, and ability to implement findings);
4. Trust in the summary and its producers (end user's perceptions of credibility markers of the work as a whole -- such as transparency, funding sources, and clear references – i.e., that the work was rigorously done by qualified individuals);

5. Quality of evidence (focused on the assessment of study quality and the totality of the evidence including how assessments were reached and information about rating); and
6. Knowledge required to understand findings (educational information that should be added to summaries due to comprehension difficulties or gaps in end user's knowledge base).

These 393 synthesized findings were then reviewed again by two authors (MKS and BC) to produce 130 recommendations for practice which, where possible, are presented based on targeted GDG members. Several recommendations also refer to particular scenarios or types of evidence syntheses such as NMA (n = 22), DTA reviews (n = 2), and updating reviews (n = 8). As previously mentioned, most studies contained diverse multidisciplinary participants. When quotes from participants were reported it was often not attributed to a specific stakeholder and several studies also included no direct quotes from participants. However, where possible, recommendations are presented according to group membership. The 130 recommendations from the qualitative synthesis are available in Figures 3, 4, and 5. Citations for recommendations can be found in Appendix 5.

A majority of recommendations related to presenting information (n = 68) or tailoring the information for the end user (n = 24). For example, items under the 'presenting information' category include things like 'use bullet points', 'flag important information by bolding/highlighting', use 'greyscale-friendly colours', and 'avoid abbreviations.' 'Tailoring information' included guidance on how to create bespoke customised documents with 'easily extractable information to forward to colleagues' and the importance of 'clarifying the audience' that the report is for and about.

Several items regarding the presentation of numerical and statistical findings were identified across several themes. For example, for 'presenting information', it was suggested to 'use absolute numbers, not probabilities' and to 'decrease numeric/statistical data' whereas the 'contextualising finding' category suggested 'interpretation aids for statistics', and noted that policy/decision makers are 'not interested in methodology.' The 'knowledge required' category highlighted the lack of awareness of abbreviations, recommending to 'avoid abbreviations (e.g., RR for relative risk, CI for confidence intervals' altogether. Some of these items are intrinsically linked as the 'knowledge required' recommendations highlighted that for readers, certain items like 'forest plots are difficult to understand' so providing 'interpretation of statistical results' and 'defining statistical terms' can be helpful.

Mixed methods synthesis

The four outcome areas for the quantitative evidence (e.g., knowledge, satisfaction) were also covered by the qualitative evidence. However, due to the large heterogeneity in stakeholders, formats, and assessments methods, it was difficult to determine whether the qualitative evidence helped explain differences in size or direction of effects in the quantitative studies.

From 74 individual quantitative findings (Appendix 4) we identified 17 which converged with at least one of the 130 qualitative recommendations (Appendix 5). Some of these 17 items supported the same recommendation (e.g., several findings supported the use of summary of findings tables) so in total these 17 quantitative findings supported 9 qualitative findings. Some of these items are inherently linked as SoF tables (4) are often using the GRADE rating scale (8). Similarly the items about assessments of quality (7 and 9) are likely to refer to GRADE as well. The 9 recommendations with mixed-methods support are marked with an asterisk in Figures 3, 4, and 5 and include providing a clear summary report that:

1. is structured,
2. is brief,
3. provides information on the standard steps and nature of the review,
4. presents results in summary of findings (SoF) tables,
5. defines statistical terms,
6. provides interpretations of statistical results,
7. includes assessments of quality,
8. describes the rating scale (GRADE), and
9. describes how authors arrived at their assessments of quality.

Throughout our recommendations, there are items which may appear at face-value to be contradictory. However, they simply accommodate different learning styles (e.g., 'use summary of findings tables' and 'use narrative summaries'), thus these are considered complimentary. Relatedly, there were some items that were expressed by different groups which echoed the end user's different needs. For example, the 'Abstract Methods Results and Discussion (AMRaD) format' was advocated by clinicians whereas 'avoid academic formatting' was expressed by policy/decision makers. Additionally there are some items that are similar but were expressed for very different purposes – for example, 'including author's names' is in both the 'presenting information' and 'trust in producers and summary' themes as some participants flagged this as a clear indicator of their trust in the quality of the work whereas others just wanted the information for general factual transparency purposes. (Appendix 5, Figures 3, 4, 5)

Discussion

This mixed methods systematic review synthesised the evidence on the effectiveness of and acceptability of different evidence synthesis summary formats. The quantitative results suggest that alternative versions of SoF tables compared to a current format and/or a standard systematic review improved knowledge or understanding. However, assessments of study quality revealed that half of the included trials had poor reporting related to the blinding of outcome assessors and those delivering treatment. While we were not able to identify a 'gold-standard' summary format, qualitative studies offered a wealth of data such that we could synthesize findings into 130 actionable recommendations across six thematic areas. To help with potential implementation, we also delineated findings by review type and stakeholder group where possible. Thirty two of the 130 recommendations were for specific types of reviews (e.g., NMA, DTA, and updating reviews). These recommendations should aid in more effective communication with different stakeholders.

The interventions included in our review were diverse with a variety of outcome measures. The majority of studies tested de novo summary prototypes making it difficult to draw comparisons. However, five studies assessed GRADE SoF tables and a significant portion of our recommendations pertain to summary of findings tables and GRADE ratings. In fact, there were enough findings concerning the quality assessment of studies and use of the GRADE scale that it warranted its own category 'quality of evidence' in the final recommendations. Previous work focused on US National Guidelines Clearinghouse clinical practice guidelines published between 2011 and 2018 found that the GRADE scale was inconsistently used and only 1 in 10 (7/67, 10.4%) guidelines explicitly reported consideration of all criteria to assess the certainty in the evidence. [61] As reflected in three of our nine recommendations with mixed-method support, GRADE is an important factor in evidence summary formats. Recent work has highlighted that there are many improvements to be made in terms of consistency in presenting GRADE symbols and explaining the recommendations. [62] This aligns with seven articles in our review which supported the need to be explicit about how the scale is used, recommending to 'provide distinct explanations of rating scale (GRADE).' Four studies also supported detailing 'how authors arrived at assessments of quality.' (Appendix 5)Many included interventions tended to be in a traditional academic style in that they were largely text based. Accordingly, numerous recommendations addressed how to 'flag important..' and 'avoid dense information' through 'structured', 'brief', and 'concise' formats with 'prominent subheadings'. The need for structured presentation of information is also supposed by previous work. Brandt et al. found that 181 internal medicine and general practice physicians had a clear preference for multi-layered guideline presentation formats. [63] Short menu formats and visual aids have been shown to improve performance when participants are presented with both conditional probability and natural frequency formats. [64] One study found that, across different levels of object numeracy and education, fact boxes (i.e., simple tabular messages) were more engaging than normal text. They also led to more comprehension and slightly more knowledge recall after six weeks compared to the same information in text. [17]

Other than MAGICApp and Tableau, no other interactive summary formats were identified in our review. Furthermore, no studies that used audio-visual strategies such as podcasts or videos were identified in this review. There is some evidence that video abstracts are more effective than graphical abstracts and traditional abstracts in comprehension, understanding, and reading experience. [65] Audio summaries also show some promising results. University staff listening to a podcast summary of a Cochrane review had the highest rates of comprehension in comparison to those who read a plain language summary or abstract. [66] Future research should explore and test these formats with GDG members.

Many general tenets were supported by multiple studies involving multidisciplinary stakeholders. For example, concerns about the presentation of numerical and statistical results resulted in recommendations across several of our themes. Similar to our findings, Cochrane's Plain Language Expectations for Authors of Cochrane Summaries (PLEACS) standards recommend presenting numerical information in terms of absolute effects and as natural frequencies. [67] A 2017 meta-analysis also supported the use of natural frequencies. Their study found that performance rates when interpreting natural frequencies increased to 24% compared to only 4% when presented in a probability format. However, three quarters of participants still failed to obtain the correct solution with either presentation. [64] On the other hand, a 2020 study by Buljan et al. found that numerical presentation (and framing) had no effect on consumer's and

biomedical student's understanding of health information in plain language summaries. [68] Previous research established that the required literacy for even plain language summaries is higher (over 10 to 15 years of education) than the recommended US 6th grade (11 or 12 years old) reading level. [69] All of this prior work reinforces the idea that effective interactions with evidence synthesis summaries requires certain baseline knowledge. This review has provided specific knowledge areas to address as detailed in the 'knowledge required' category (e.g., the need to define terms, explain methodologies, grading scales, and statistics, and generally provide a supplemental explanation sheet to end users). Initiatives such as the International Guideline Development Credentialing and Certification Program (INGUIDE) [71] may also help address some of these knowledge needs by ensuring that guideline development group members have the necessary competencies.

Our recommendations are proposals for consideration, not strict rules for practice, especially considering that the evidence-base supporting many recommendations is weak and not all may be practical for resource-limited teams. Included studies often did not discuss time or resources required to actually produce the summary format(s) which could make implementation difficult, especially in light of the large number (n = 130) of recommendations. Inclusion of certain items, particularly those related to 'contextualising' findings may require additional work or expertise which some may consider to be outside the scope of a typical review. [58] However, these suggestions should not be ignored as research has shown that context is rarely provided in sufficient detail in existing reviews and guidelines [72]and applying evidence synthesis findings to local contexts is a major weakness reported by some health technology assessment (HTA) units trying to promote healthcare decision-making. [73]

The strengths of this study include the mixed methods approach and an extensive search strategy. However, our study has several limitations. Firstly, we did not include observational studies although during screening we excluded few studies based on their study design (Fig. 2) [2]. The main limitations of our findings relates to the issues of completeness of the reporting of included studies. Several articles did not provide a copy or access to the summary format(s) tested so it was sometimes difficult to properly contextualise their results. Additionally, it was often difficult to attribute a finding to a specific stakeholder group as included studies often did not provide group membership details about quotes used. This meant that many of our recommendations are non-specific as we were unable to fully decipher what works for who and under which circumstances. Stakeholders involved in guideline development have different styles of reasoning and knowledge bases to draw from [74], therefore drawing conclusions that are stakeholder group specific is complex. Even within one group (e.g., patient representatives), one size does not fit all when presenting recommendations. [70] However, we recommend that future work with multidisciplinary stakeholders should denote group membership when reporting quotes from participants as this was a deficit in our included studies. For example, while there is some reporting guidance for what public or patient version of clinical guidelines should include [75], we are still missing a step in the process wherein it is unclear what works best for patient representatives involved in clinical guideline development groups. Lastly, we excluded studies in the general population and students. Studies have shown that PLS improved understanding in these populations. [76, 77]

Conclusions

Our results provide valuable information that can be used to improve existing formats and inform future research aimed at developing more effective evidence synthesis summary formats. Future research should further explore these proposed recommendations amongst the different guideline development group members to further explore which items are particularly important for which stakeholder. Our research team plans to conduct a prioritisation exercise for these recommendations so we can use them as guidance for focus group workshops with GDG members. Furthermore, other mediums of summary formats not identified in this review could be explored further such as the use of podcasts or video abstracts or summaries.

References

1. Clyne B, Sharp M. Evidence synthesis and translation of findings for national clinical guideline development: addressing the needs and preferences of guideline development groups. OSF; 2021 [cited 2021 Jun 3]; Available from: https://osf.io/sk4nx/.
2. Sharp M, Clyne B. Evidence synthesis summary formats for decision-makers and Clinical Guideline Development Groups: A mixed-methods systematic review protocol. OSF; 2021 [cited 2021 Jul 7]; Available from: https://osf.io/jdauy.
3. Eccles MP, Grimshaw JM, Shekelle P, Schünemann HJ, Woolf S. Developing clinical practice guidelines: target audiences, identifying topics for guidelines, guideline group composition and functioning and conflicts of interest. Implement Sci. 2012;7:60.
4. Woolf S, Schünemann HJ, Eccles MP, Grimshaw JM, Shekelle P. Developing clinical practice guidelines: types of evidence and outcomes; values and economics, synthesis, grading, and presentation and deriving recommendations. Implement Sci. 2012;7:61.

5. Qaseem A, Forland F, Macbeth F, Ollenschläger G, Phillips S, van der Wees P. Guidelines International Network: Toward International Standards for Clinical Practice Guidelines. Ann Intern Med American College of Physicians. 2012;156:525–31.
6. Wieringa S, Engebretsen E, Heggen K, Greenhalgh T. Clinical guidelines and the pursuit of reducing epistemic uncertainty. An ethnographic study of guideline development panels in three countries. Soc Sci Med. 2021;272:113702.
7. Abdelhamid A, Howe A, Stokes T, Qureshi N, Steel N. Primary care evidence in clinical guidelines: a mixed methods study of practitioners' views. Br J Gen Pract. 2014;64:e719–27.
8. Correa VC, Lugo-Agudelo LH, Aguirre-Acevedo DC, Contreras JAP, Borrero AMP, Patiño-Lugo DF, et al. Individual, health system, and contextual barriers and facilitators for the implementation of clinical practice guidelines: a systematic metareview. Health Res Policy Syst. 2020;18:74.
9. Dobbins MCR, Barnsley J. Factors affecting the utilization of systematic reviews. A study of public health decision makers. 2001;17:203–14.
10. Alonso-Coello P, Schünemann HJ, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. BMJ [Internet]. British Medical Journal Publishing Group; 2016 [cited 2020 Jul 7];353. Available from: https://www.bmj.com/content/353/bmj.i2016.
11. Zhang Y, Alonso-Coello P, Guyatt Gh Y-N, Jj A, Ea, Hazlewood G, et al. GRADE Guidelines: 19. Assessing the certainty of evidence in the importance of outcomes or values and preferences-Risk of bias and indirectness. J Clin Epidemiol. 2018;111:94–104.
12. Wallace JW; B Charles; Clarke, Mike. Improving the uptake of systematic reviews: a systematic review of intervention effectiveness and relevance. 2014;4:e005834-NA.
13. Tricco AC; C Thomas R, Motiwala SM, Sullivan SS, Kealey S, Hemmelgarn MR, Ouimet BR, Hillmer M, Perrier MP, Shepperd L, Straus S. Sharon E. Barriers and facilitators to uptake of systematic reviews by policy makers and health care managers: a scoping review. 2016;11:4–4.
14. Perrier L, Mrklas K, Lavis JN, Straus SE. Interventions encouraging the use of systematic reviews by health policymakers and managers: A systematic review. Implement Sci. 2011;6:43.
15. Petkovic J, Welch V, Jacob MH, Yoganathan M, Ayala AP, Cunningham H, et al. The effectiveness of evidence summaries on health policymakers and health system managers use of evidence from systematic reviews: a systematic review. Implement Sci. 2016;11:162.
16. Chambers DWP, Thompson C, Hanbury A, Farley K, Light K. Maximizing the impact of systematic reviews in health care decision making: a systematic scoping review of knowledge-translation resources. 2011;89:131–56.
17. Brick C, McDowell M, Freeman ALJ. Risk communication in tables versus text: a registered report randomized trial on 'fact boxes'. Royal Society Open Science. Royal Society; 7:190876.
18. Bressan V, Bagnasco A, Aleo G, Timmins F, Barisone M, Bianchi M, et al. Mixed-methods research in nursing – a critical review. J Clin Nurs. 2017;26:2878–90.
19. Lizarondo L, Stern C, Carrier J, Godfrey C, Rieger K, Salmond S, et al. Chapter 8: Mixed Methods Systematic Reviews. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis [Internet]. Joanna Briggs Institute; 2020 [cited 2021 Feb 23]. Available from: https://wiki.jbi.global/display/MANUAL/Chapter+8%3A+Mixed+methods+systematic+reviews.
20. Lizarondo L, Stern C, Apostolo J, Carrier J, de Borges K, Godfrey C, et al. Five common pitfalls in mixed methods systematic reviews – lessons learned. Journal of Clinical Epidemiology. 2022;S0895435622000750.
21. Sharp MK, Tyner B, Baki DABA, Farrell C, Devane D, Mahtani KR, et al. Evidence synthesis summary formats for clinical guideline development group members: a mixed-methods systematic review protocol [Internet]. HRB Open Research; 2021 [cited 2021 Aug 23]. Available from: https://hrbopenresearch.org/articles/4-76.
22. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. Journal of Clinical Epidemiology. Elsevier; 2021.
23. Covidence. - Better systematic review management [Internet]. Covidence. [cited 2021 Feb 15]. Available from: https://www.covidence.org/.
24. Haddaway NR. citationchaser: an R package for forward and backward citations chasing in academic searching [Internet]. Zenodo; 2021 [cited 2021 Feb 15]. Available from: https://zenodo.org/record/4533747.
25. nealhaddaway. nealhaddaway/citationchaser [Internet]. 2021 [cited 2021 Feb 15]. Available from: https://github.com/nealhaddaway/citationchaser.

26. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ. Br Med J Publishing Group. 2014;348:g1687.
27. critical-appraisal-tools. - Critical Appraisal Tools | Joanna Briggs Institute [Internet]. [cited 2021 Feb 24]. Available from: https://jbi.global/critical-appraisal-tools.
28. Harbord RM, Egger M, Sterne JAC. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. Stat Med. 2006;25:3443–57.
29. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ Br Med J Publishing Group. 1997;315:629–34.
30. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions [Internet]. Cochrane; 2020 [cited 2021 Feb 23]. Available from: .
31. Lockwood C, Munn Z, Porritt K. Qualitative research synthesis: methodological guidance for systematic reviewers utilizing meta-aggregation. JBI Evid Implement. 2015;13:179–87.
32. Tufanaru C. Theoretical Foundations of Meta-Aggregation: Insights from Husserlian Phenomenology and American Pragmatism [Internet]. [Adelaide, Australia]: The Joanna Briggs Institute. The University of Adelaide; 2015. Available from: https://digital.library.adelaide.edu.au/dspace/bitstream/2440/98255/1/01front.pdf.
33. Hannes K, Lockwood C. Pragmatism as the philosophical foundation for the Joanna Briggs meta-aggregative approach to qualitative evidence synthesis. J Adv Nurs. 2011;67:1632–42.
34. NVivo. Qualitative Data Analysis Software | NVivo [Internet]. 2021 [cited 2020 Sep 9]. Available from: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home.
35. 4.3.4.5.3 Textual data synthesis - JBI Manual for Evidence Synthesis - JBI Global Wiki [Internet]. [cited 2021 Dec 8]. Available from: https://jbi-global-wiki.refined.site/space/MANUAL/3283911241/4.3.4.5.3%20Textual%20data%20synthesis.
36. Hong QN, Pluye P, Bujold M, Wassef M. Convergent and sequential synthesis designs: implications for conducting and reporting systematic reviews of qualitative and quantitative evidence. Syst Reviews. 2017;6:61.
37. Sandelowski M, Voils CI, Barroso J. Defining and Designing Mixed Research Synthesis Studies. Res Sch. 2006;13:29.
38. Babatunde OO, Tan V, Jordan JL, Dziedzic K, Chew-Graham CA, Jinks C, et al. Evidence flowers: An innovative, visual method of presenting “best evidence” summaries to health professional and lay audiences. Res Synthesis Methods. 2018;9:273–84.
39. Buljan I, Tokalić R, Roguljić M, Zakarija-Grković I, Vrdoljak D, Milić P, et al. Comparison of blogshots with plain language summaries of Cochrane systematic reviews: a qualitative study and randomized trial. Trials. 2020;21:426.
40. Dobbins M, DeCorby K, Twiddy T. A knowledge transfer strategy for public health decision makers. Worldviews on Evidence-Based Nursing. 1. Malden: Wiley-Blackwell; 2004. pp. 120–8.
41. Hartling L, Guise J-M, Hempel S, Featherstone R, Mitchell MD, Motu’apuaka ML, et al. EPC Methods: AHRQ End-User Perspectives of Rapid Reviews. Agency for Healthcare Research and Quality (US); 2016; Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medp&NEWS=N&AN=27195347.
42. Hartling L, Guise JM, Hempel S, Featherstone R, Mitchell MD, Motu’apuaka ML, et al. Fit for purpose: Perspectives on rapid reviews from end-user interviews. Systematic Reviews [Internet]. 2017;6. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85013130994&doi=10.1186%2fs13643-017-0425-7&partnerID=40&md5=d0ea651ce13e9b75e702f2b5a9e822cc.
43. Marquez C, Johnson AM, Jassemi S, Park J, Moore JE, Blaine C, et al. Enhancing the uptake of systematic reviews of effects: what is the best format for health care managers and policy-makers? A mixed-methods study. Implementation Science [Internet]. 2018;13. Available from:://WOS:000436144000001.
44. Mustafa RA, Wiercioch W, Santesso N, Cheung A, Prediger B, Baldeh T, et al. Decision-Making about Healthcare Related Tests and Diagnostic Strategies: User Testing of GRADE Evidence Tables. PLoS ONE [Internet]. 2015;10. Available from:://WOS:000363185500001.
45. Newberry SJ, Shekelle PG, Vaiana M, Motala A. Reporting the Findings of Updated Systematic Reviews of Comparative Effectiveness: How Do Users Want To View New Information? Agency for Healthcare Research and Quality (US); 2013; Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medp&NEWS=N&AN=23785728.
46. Opiyo N, Shepperd S, Musila N, Allen E, Nyamai R, Fretheim A, et al. Comparison of Alternative Evidence Summary and Presentation Formats in Clinical Guideline Development: A Mixed-Method Study. PLoS ONE [Internet]. 2013;8. Available from:://WOS:000315210400056.

47. Rosenbaum SGC, Nylund HK, Oxman AD. User testing and stakeholder feedback contributed to the development of understandable and useful Summary of Findings tables for Cochrane reviews. 2010;63:607–19.
48. Smith CJ, Jungbauer RM, Totten AM. Visual Evidence: Increasing Usability of Systematic Reviews in Health Systems Guidelines Development. Appl Clin Inf. 2019;10:743–50.
49. Totten AM, Smith C, Dunham K, Jungbauer RM, Graham E. Improving Access to and Usability of Systematic Review Data for Health Systems Guidelines Development. Agency for Healthcare Research and Quality (US); 2019; Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medp&NEWS=N&AN=31013017.
50. Buljan I, Malički M, Wager E, Puljak L, Hren D, Kellie F, et al. No difference in knowledge obtained from infographic or plain language summary of a Cochrane systematic review: three randomized controlled trials. J Clin Epidemiol. 2018;97:86–94.
51. Carrasco-Labra A, Brignardello-Petersen R, Santesso N, Neumann I, Mustafa RA, Mbuagbaw L, et al. Improving GRADE evidence tables part 1: a randomized trial shows improved understanding of content in summary of findings tables with a new format. J Clin Epidemiol Elsevier. 2016;74:7–18.
52. Rosenbaum SGC, Oxman AD. Summary-of-findings tables in Cochrane reviews improved understanding and rapid retrieval of key information. 2010;63:620–6.
53. Steele R. Mental health clinicians views of summary and systematic review utility in evidence-based practice. Health Information and Libraries Journal [Internet]. Available from::://WOS:000627057300001.
54. Perrier LKM, Ryan; Straus SE. An iterative evaluation of two shortened systematic review formats for clinicians: a focus group study. 2014;21:e341-6.
55. Perrier LKM, Ryan; Straus SE. A usability study of two formats of a shortened systematic review for clinicians. 2014;4:e005919-NA.
56. Busert LK, Mütsch M, Kien C, Flatz A, Griebler U, Wildner M, et al. Facilitating evidence uptake: Development and user testing of a systematic review summary format to inform public health decision-making in German-speaking countries. Health Research Policy and Systems [Internet]. 2018;16. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85049782278&doi=10.1186%2fs12961-018-0307-z&partnerID=40&md5=8a60b2081f09fd2655dac0ddecb23467.
57. Hartling L, Gates A, Pillay J, Nuspl M, Newton AS. Development and Usability Testing of EPC Evidence Review Dissemination Summaries for Health Systems Decisionmakers. Agency for Healthcare Research and Quality (US); 2018; Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medp&NEWS=N&AN=30507111.
58. Rosenbaum SE, Glenton C, Wiysonge CS, Abalos E, Mignini L, Young T, et al. Evidence summaries tailored to health policy-makers in low- and middle-income countries. Bull World Health Organ. 2011;89:54–61.
59. Yepes-Nunez JJ, Li SA, Guyatt G, Jack SM, Brozek JL, Beyene J, et al. Development of the summary of findings table for network meta-analysis. J Clin Epidemiol. 2019;115:1–13.
60. Mustafa R, Wiercioch W, Brozek J, Lelgemann M, Buehler D, Garg A, et al. Enhancing the acceptance and implementation of grade summary tables for evidence about diagnostic tests. BMJ Qual Saf. 2013;22:A36.
61. Dixon C, Dixon PE, Sultan S, Mustafa R, Morgan RL, Murad MH, et al. Guideline developers in the United States were inconsistent in applying criteria for appropriate Grading of Recommendations, Assessment, Development and Evaluation use. J Clin Epidemiol. 2020;124:193–9.
62. Klugar M, Kantorová L, Pokorná A, Líčeník R, Dušek L, Schünemann HJ, et al. Visual transformation for guidelines presentation of the strength of recommendations and the certainty of evidence. J Clin Epidemiol Elsevier. 2022;143:178–85.
63. Brandt L, Vandvik PO, Alonso-Coello P, Akl EA, Thornton J, Rigau D, et al. Multilayered and digitally structured presentation formats of trustworthy recommendations: A combined survey and randomised trial. BMJ Open [Internet]. 2017;7. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85012273660&doi=10.1136%2fbmjopen-2016-011569&partnerID=40&md5=afb3d847be40dd8ac9308fb00601d1ee.
64. Meta-analysis of the effect of natural frequencies on Bayesian reasoning. - PsycNET [Internet]. [cited 2021 Jul 14]. Available from: https://content.apa.org/record/2017-47164-001.
65. Bredbenner K, Simon SM. Video abstracts and plain language summaries are more effective than graphical abstracts and published abstracts. PLoS ONE. 2019;14:e0224697.
66. Maguire LC, Mike. How much do you need: a randomised experiment of whether readers can understand the key messages from summaries of Cochrane Reviews without reading the full review. 2014;107:444–9.
67. Plain Language Expectations for Authors of Cochrane Summaries (PLEACS). - Cochrane Editorial and Publishing Policy Resource - Confluence [Internet]. [cited 2021 Jul 12]. Available from: https://documentation.cochrane.org/pages/viewpage.action?

pageId=117380534.
68. Buljan I, Tokalic R, Roguljic M, Zakarija-Grkovic I, Vrdoljak D, Milic P, et al. Framing the numerical findings of Cochrane plain language summaries: two randomized controlled trials. BMC Medical Research Methodology [Internet]. 2020;20. Available from::://WOS:000533890400002.
69. Karačić JDP, Buljan I, Hren D, Marušić. Ana. Languages for different health information readers: multitrait-multimethod content analysis of Cochrane systematic reviews textual summary formats. 2019;19:1–9.
70. Fearns N, Walker L, Graham K, Gibb N, Service D. User testing of a Scottish Intercollegiate Guideline Network public guideline for the parents of children with autism. BMC Health Serv Res. 2022;22:77.
71. International Guideline Credentialing &. Certification Program [Internet]. [cited 2022 Mar 31]. Available from: https://inguide.org/.
72. Booth A, Moore G, Flemming K, Garside R, Rollins N, Tunçalp Ö, et al. Taking account of context in systematic reviews and guidelines considering a complexity perspective. BMJ Global Health BMJ Specialist Journals. 2019;4:e000840.
73. Poder TG, Rhainds M, Bellemare CA, Deblois S, Hammana I, Safianyk C, et al. Experiences of Using Cochrane Systematic Reviews by Local HTA Units. International Journal of Health Policy and Management [Internet]. Iran; 2020; Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medp&NEWS=N&AN=32772006.
74. Wieringa S, Dreesens D, Forland F, Hulshof C, Lukersmith S, Macbeth F, et al. Different knowledge, different styles of reasoning: a challenge for guideline development. BMJ Evid Based Med. 2018;23:87–91.
75. Wang X, Chen Y, Akl EA, Tokalić R, Marušić A, Qaseem A, et al. The reporting checklist for public versions of guidelines: RIGHT-PVG. Implement Sci. 2021;16:10.
76. Santesso NRT, Nilsen ES, Glenton C, Rosenbaum S, Ciapponi A, Moja L, Pardo JP, Zhou Q, Schünemann H. J. A summary to communicate evidence from systematic reviews to the public improved understanding and accessibility of information: a randomized controlled trial. 2014;68:182–90.
77. Alderdice FMJ, Lasserson TJ, Beller E, Carroll M, Hundley V, Sunderland J, Devane D, Noyes J, Key S, Norris S, Wyn-Davies J, Clarke. Mike. Do Cochrane summaries help student midwives understand the findings of Cochrane systematic reviews: the BRIEF randomised trial. 2016;5:40–40.

Declarations

Ethics approval and consent to participate

There were no human participants involved in this project so ethical approval is not necessary. All data is publicly available.

Consent for publication

Not applicable.

Availability of data and materials

The study was previously pre-registered on Open Science Framework and the protocol was published in HRB Open Research. [1,2] The datasets generated and/or analysed during the current study are available on Open Science Framework (OSF). Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Competing interests

The authors declare that they have no other competing interests.

Funding

MKS and BC are supported by a Health Research Board (HRB) Emerging Investigator Award (EIA-2019-09). The HRB is not involved in the project's protocol, analysis plan, data collection, analysis, or interpretation of study results.

Authors' contributions

Study conceptualisation: BC, MON, MR, SMS. **Methodology development:** MKS, DD, KRM, SMS, MON, MR, BC. **Data curation:** MKS, DAB, JQ, BT, BC. **Formal analysis:** MKS, BC. **Writing-reviewing and editing:** MKS, DAB, JQ, BT, DD, KRM, SMS, MO, MR, BC. **Supervision:** MKS, BC. **Project administration:** MKS, BC. **Funding acquisition:** BC, MON, MR, SMS, KRM.

Tables

Table 1. Included Qualitative Studies

Author *(Year, Country)*	Collection Method	Participants	Format(s) evaluated	Topics
Babatunde *(2018, England)* [38]	Semi-structured interviews *(and questionnaires)* [b]	N = 21. Clinicians (11), researchers (5), epidemiologists (3), health service/trial managers (2)	Evidence flowers and summary table	Musculoskeletal conditions
Buljan *(2020, Croatia)* [39]	Focus groups	N = 20. Patient advocates (9), doctors (4), *Medical students (7)* [a]	Plain language summary, infographic, scientific abstract	Breech presentation
Busert *(2018, International)* [56]	Semi-structured interviews	N = 18. Public-health decision makers	4 page summary with Summary of Findings (SoF) table and Grading of Recommendations, Assessment, Development and Evaluations (GRADE) ratings	Food, alcohol, and tobacco portion/packaging
Dobbins *(2004, Canada)* [40]	Focus groups	N = 46. Medical officers (7), program managers/coordinators (25), decision makers (14)	Summary statement	Tobacco control
Hartling *(2016, 2017, USA)* [41,42]	Semi-structured interviews	N = 8. Guideline developers (3), healthcare providers (3), research funders (1), health insurers (1)	'Rapid products' (evidence inventory, rapid response, rapid review)	Venous thromboembolism
Hartling *(2018, Canada)* [57]	Semi-structured interviews	N = 6. Decision makers	3-page summary	Youth mental health
Marquez *(2018, Canada)*[43]	Semi-structured interviews *(and survey)* [b] Semi-structured interviews	N = 11. Health care managers (5), policymakers (6) N = 12. Health care managers (5), policymakers (7)	Summary prototype	Healthcare management/ services
Mustafa *(2015, International)* [44]	Semi-structured interviews, workshop discussions	N = 20. Researchers, health professionals, guideline developers	3 formats[c] of GRADE evidence tables	Diagnostic test accuracy reviews
Newbery *(2013, USA)* [45]	Focus groups *(with questionnaires)* Individual feedback *(and questionnaires)*	N = 15. Health insurer (2), insurance/former policymaker (2), clinicians (3), researchers (2), governmental research directors (2), research consultant (1) N= 3. Community physicians	7 differently formatted executive summaries	Acute otis media
Opiyo *(2013, Kenya)* [46]	Semi-structured interviews	N = 16. Multidisciplinary guideline development group members	SoF tables, graded-entry summary, normal systematic review	Newborn care, hand hygiene
Perrier *(2014, Canada)* [55]	Focus groups	N = 10. Family physicians	Case-based and evidence-based prototypes	Rosacea
Perrier *(2014, Canada)* [54]	Focus groups	N = 32. Primary care physicians	Two summary prototypes	Rosacea
Rosenbaum *(2011, International)* [58]	Semi-structured interviews	N = 18. Policymakers and managers	Short summaries	Healthcare management/ services
Rosenbaum *(2010, International)* [47]	Semi-structured interviews (and workshops)	N = 21. Health professionals, researchers	SoF tables	Deep vein thrombosis

Smith, Totten *(2019, USA)* [48,49]	Semi-structured interviews	N = 6. Department director (1), health system experts (4), guideline developers (2)	MAGICapp, Tableau	Chronic pain
Steele *(2021, England)* [53]	Semi-structured interviews	N = 7. Mental health clinicians	One-page summary, full systematic review	Mental health
Yepes-Nunez *(2019, International)* [59]	Semi-structured interviews	N = 32. Methodologists (21), meta-analysis users (5), clinicians (6)	SoF tables	Network meta analyses

[a] Population and accompanying RCT not eligible. [b] Not eligible [c] There were 4 formats total but only 3 were shown to the user testing group.

Abbreviations: United States of America (USA), Germany, Austria, and Switzerland (DACH), Summary of findings (SoF), Grading of Recommendations, Assessment, Development and Evaluations (GRADE), MAGIC(Making GRADE the Irresistible Choice)

Table 2. Included Randomised Controlled Trials

Author (Year, Country)	**Participants**	**Intervention and Comparators**	**Primary *(Secondary)* Outcomes and Operationalization (Number of questions, type, scales)**	**Focus**
Buljan (2018, Croatia) [50]	N = 163 (eligible across trials) 99 patient representatives, 64 doctors *(171 students)*[a]	Infographic, plain language summary, scientific abstract (doctors only)	Understanding/knowledge (10, open ended) *Reading experience (5, summative, 10-point scale)* *User-friendliness (5, summative, 10-point scale)*	Breech presentation
Carrasco-Labra (2016, International) [51]	N = 284 Health professionals (122), guideline developers (42), researchers (120)	2 versions (1 existing, 1 alternate) of Grading of Recommendations, Assessment, Development and Evaluations (GRADE) Summary of Findings (SoF) tables	Understanding (7, multiple choice, 5-point *scale*) *Accessibility of information (3, 7-point scale; 1, 5-point scale)* *Satisfaction (6, yes/no)* *Preference (1, 7-point scale)*	Paediatric probiotics
Opiyo (2013, Kenya) [46]	N = 70 Paediatricians (32), medical/nursing officers (18), researchers (5), healthcare trainers (5), governmental/clinical officers (7), pharmacists (2), administrator (1)	3 different topic 'evidence packs' 1. Normal systematic review (SR) 2. SR plus SoF tables 3. Graded-entry SR	Understanding (2 per format, 3-point scale) *Composite endpoint (1, 5-point scale): Clarity (1 per format, 3-point scale)* *Accessibility (2 per format, 5-point scale)*	Hand hygiene, newborn care, newborn feeding regimens
Rosenbaum (2010, International) [52]	N = 72 (RCT1) Health care professionals N = 33 (RCT2) Staff from Cochrane entities	Normal Cochrane review (CR) with no SoF table, CR with SoF table (limited formatting), CR with SoF table (full formatting) Normal Cochrane review (CR) with no SoF table, CR with SoF table (revised)	User satisfaction (unclear, multiple choice), Perceived understanding and ease of use *(7, 8-point scale)* Understanding (4, unclear), Time spent finding key results (1, continuous)	Deep vein thrombosis

[a] Population does not meet eligibility criteria for this review. [c]65 participants completed the questionnaires. Group membership details are given for these 65, not the full 70 enrolled in the study.

Abbreviations: Summary of findings (SoF), Grading of Recommendations, Assessment, Development and Evaluations (GRADE), Cochrane Review (CR)

Table 3. Description of interventions in randomised controlled trials

Author (Year)	Brief Description of Intervention	Intervention Location, Mode of Delivery, Time limit	Materials and components
Buljan (2018, Croatia) [50]	Infographic, plain language summary, scientific abstract	Online, Electronic, None	Participants read one of the summary formats, followed by the survey (first a numeracy test with sufficient delay for the knowledge test). Patient representatives were presented with the infographic or PLS and doctors were presented all three formats.
Carrasco-Labra (2016, International) [51]	Summary of findings table	Online, Electronic, 25 minutes	Participants were exposed to one table containing either the new or current format and the outcomes understanding, accessibility of information, satisfaction, and preference were assessed. Participants were then shown the table to which they were not initially allocated and their preference was assessed.
Opiyo (2013, Kenya) [46]	Evidence summaries in 3 formats (A, B, C)	In person workshop, Paper, 45 minutes	Summaries were delivered to participants as pre-reading materials one month before the workshop. Participants completed questionnaires on the first day of the guideline development workshop before the panel discussions about guidance recommendations.
Rosenbaum (2010, International) [52]	RCT1: Normal Cochrane review (CR) with no SoF table, CR with SoF table (limited formatting), CR with SoF table (full formatting) RCT2: Normal Cochrane review (CR) with no SoF table, CR with SoF table (revised)	RCT1: In person workshop, Unclear, Unclear RCT2: In person workshop, Unclear, Unclear	Participants first answered a questionnaire based on the version of the review they had received. Then all participants were shown both formatting versions of the SoF tables and were instructed to answer a final set of questions measuring their preferences and attitudes about the inclusion Summary of Findings table in reviews.

Abbreviations: Summary of findings (SoF), Cochrane Review (CR)

a

Table 4. Quantitative Results

Author (Year) Interventions	Primary *(Secondary)* Outcome Measures	Results			
		Understanding/ Knowledge	Satisfaction/ Reading Experience	Accessibility/ Ease of use	Preference
Buljan (2018) [50] - Infographic - PLS - SA (Doctors only)	Understanding/knowledge (max score = 10) *Reading experience (max score = 50)* *User-friendliness (*max score = 50*)*	Patients (n=99), median score (95% CI) Infographic: 7.0 (6.0–7.0) PLS: 7.0 (6.0–7.0) P= 0.511 Doctors (n=64), median score (95% CI) Infographic: 8.0 (6.0–8.0) PLS: 8.0 (7.0–9.0) SA: 8.0 (5.9–9.0) P= 0.611 Significant predictor of knowledge score: 1. Patients only: awareness of Cochrane SRs (OR 5.3; 95% CI: 1.7 - 16.6), 13.4% of variance	*Reading experience* Patients (n=99), median score (95% CI) Infographic: 33.0 (28.0–36.0) PLS: 22.5 (19.0–27.4) P<0.001 Doctors (n=64), median score (95% CI) Infographic: 37.0 (26.8–41.3) PLS: 32.0 (30.0–39.9) SA: 24.0 (21.3–27.2) P= 0.002	*User-friendliness* Patients (n=99), median score (95% CI) Infographic: 30.0 (25.5–34.5) PLS: 21.0 (19.0–25.0) P<0.001 Doctors (n=64), median score (95% CI) Infographic: 36.0 (30.9–40.0) PLS: 29.0 (26.8–36.2) SA: 25.0 (23.5–27.2) P= 0.003	Not reported
Carrasco-Labra (2016) [51] - Existing GRADE SoF table - Alternate GRADE SoF tables	Understanding (7 multiple choice questions on 5-point scale, analysed at question level), *Accessibility of information (5-point scale),* *Satisfaction (6 yes/no questions analysed at question level),* *Preference (7-point scale)*	4/7 items risk difference (RD, 95% CI) in favour of alternate SoF tables 1. Understanding of quality of evidence and treatment effect RD: 62% (52 - 71) p <0.001 2. Ability to determine risk difference RD: 63% (54.6 - 71) p <0.001 3. Ability to quantify risk RD: 6% (0.1 - 13.3) p= 0.06 4. Understanding of quality of evidence RD: 7% (0.1 - 12.4) p= 0.06 3/7 items similar results (RD 95% CI)	Questions where largest proportion in favour of alternate SoF tables: 5/6 Questions where largest proportion in favour of existing SoF table: 1/6	Overall accessibility mean difference (MD (SE)) in favour of alternate SoF: MD 0.3 (0.11) P=0.001	MD (SE) in favour of alternate SoF: 2.8 (1.6)

		between formats: 1. Ability to interpret risk RD: 0% (-5.3 - 5.4), p=0.99 2. Ability to relate N of participant/studies and outcomes RD: -3% (-7.5 - 1.7) p= 1.00 3. Ability to interpret footnotes RD: 7% (-2 - 15), p=0.18			
Opiyo (2013) [46] - Normal systematic review (SR) - SR plus SoF tables - Graded-entry SR	Understanding (2 questions per format on 3-point scale) *Composite endpoint (on 5-point scale):* *Clarity (1 question per format on 3-point scale)* *Accessibility (2 questions per format on 5-point scale)*	Odds Ratio (OR) (95% CI) SR plus SoF versus SR, OR 0.59 (0.32 - 1.07) Graded-entry SR versus SR, OR 0.66 (0.36 - 1.21) Sub-group analyses: Policy makers understanding SR plus SoF OR 1.5 (0.15 - 15.15) Graded- entry OR 1.5 (0.64 t-3.54)	Not reported	*Accessibility* SR plus SoF versus SR: - OR (95% CI) 0.91 (0.57 - 1.46) - MD (95% CI) 0.11 (-0.71 – 0.48) Graded-entry SR versus SR: - OR 1.06 (1.06 to 2.20) - MD (95% CI) 0.52 (0.06-0.99)	Not reported
Rosenbaum (2010) [52] RCT 1 - Cochrane review (CR) with no SoF table - CR with SoF table (limited formatting), - CR with SoF table (full formatting)	User satisfaction (multiple choice questionnaire), Perceived understanding and ease of use *(7 questions on 8-point scale)*	Proportion who agree/strongly agree main findings were easy to understand % (95% CI) No SoF table: 56 (37-75) With SoF table (both formats): 60 (46-74) p=0.54	Not reported	Proportion who agree/strongly agree very accessible % (95% CI) No SoF table: 17 (2-32) With SoF table (both formats): 41 (27-56) p=0.037	65% agreed CR should include SOF with the proposed format
Rosenbaum (2010) RCT 2 - Cochrane review (CR) with no SoF table - CR with SoF table	Understanding (4 questions) Time spent finding 5 key results	2/4 difference in proportion of questions correctly answer (%, 95% CI) favouring with SOF:	Not reported	Mean time (min) finding answers Risk in the control group: No SoF table: 4 With SoF table: 1.5 p=0.02	84% agreed CR should include SOF with the proposed format

(limited formatting), - CR with SoF table (full formatting)		1. risk in the control group: 44% (21–67) versus 93% (81–100), p=0.003) 2. risk in the intervention group: 11% (0–26) versus 87% (69–100), p<0.001) 2/4 no difference in proportion of questions correctly answer (%, 95% CI): 1. Confidence of review authors: without SoF 67% (45–88) vs. with SoF 87% (69–100), P = .18 2. Identifying important outcomes: without SoF 33% (9–57) versus with SoF 53% (28–79), p = .27		Risk in the intervention group: No SoF table: 2.8 With SoF table: 1.3 p=0.118 Confidence of review authors: No SoF table: 1.5 With SoF table: 2.1 p=0.47 Identifying important outcomes: No SoF table: 1.9 With SoF table: 2.0 p=0.88	
Abbreviations: CI: Confidence Interval, CR: Cochrane Review, GRADE: Grading of Recommendations, Assessment, Development and Evaluations, MD: Mean Difference, NR: Not Reported, OR: Odds Ration, PLS: Plain language summary, RD: Risk Difference, SA: Scientific Abstract, SD: Standard Deviation, SE: Standard Error, SoF: Summary of findings, SR: Systematic Review					

Figures

Figure 1. Mixed Methods Synthesis Steps and Results

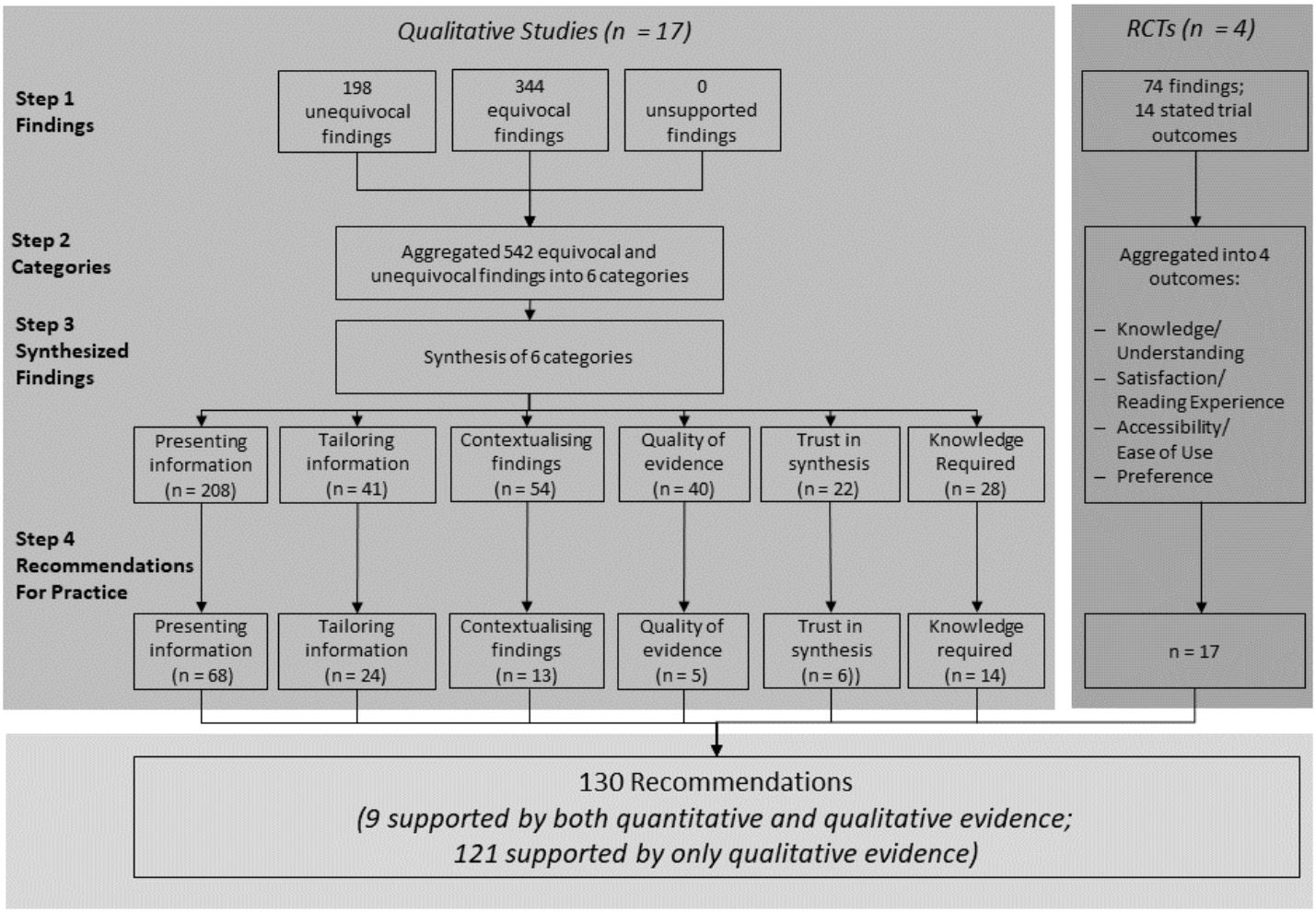
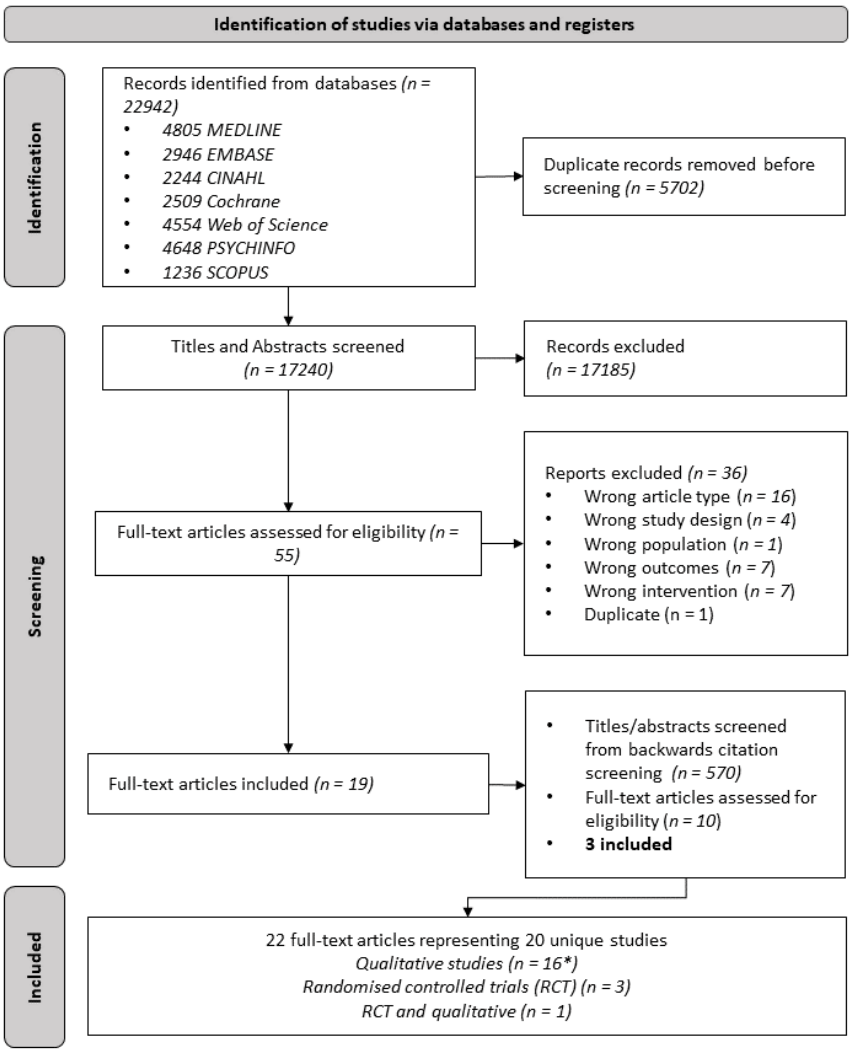


Figure 1

Mixed Methods Synthesis Steps and Results

Figure 2. PRISMA Flow Diagram



** One article included both a qualitative study and RCT but the population of the RCT was not eligible*

Figure 2

PRISMA Flow Diagram

Figure 3. Recommendations for Practice

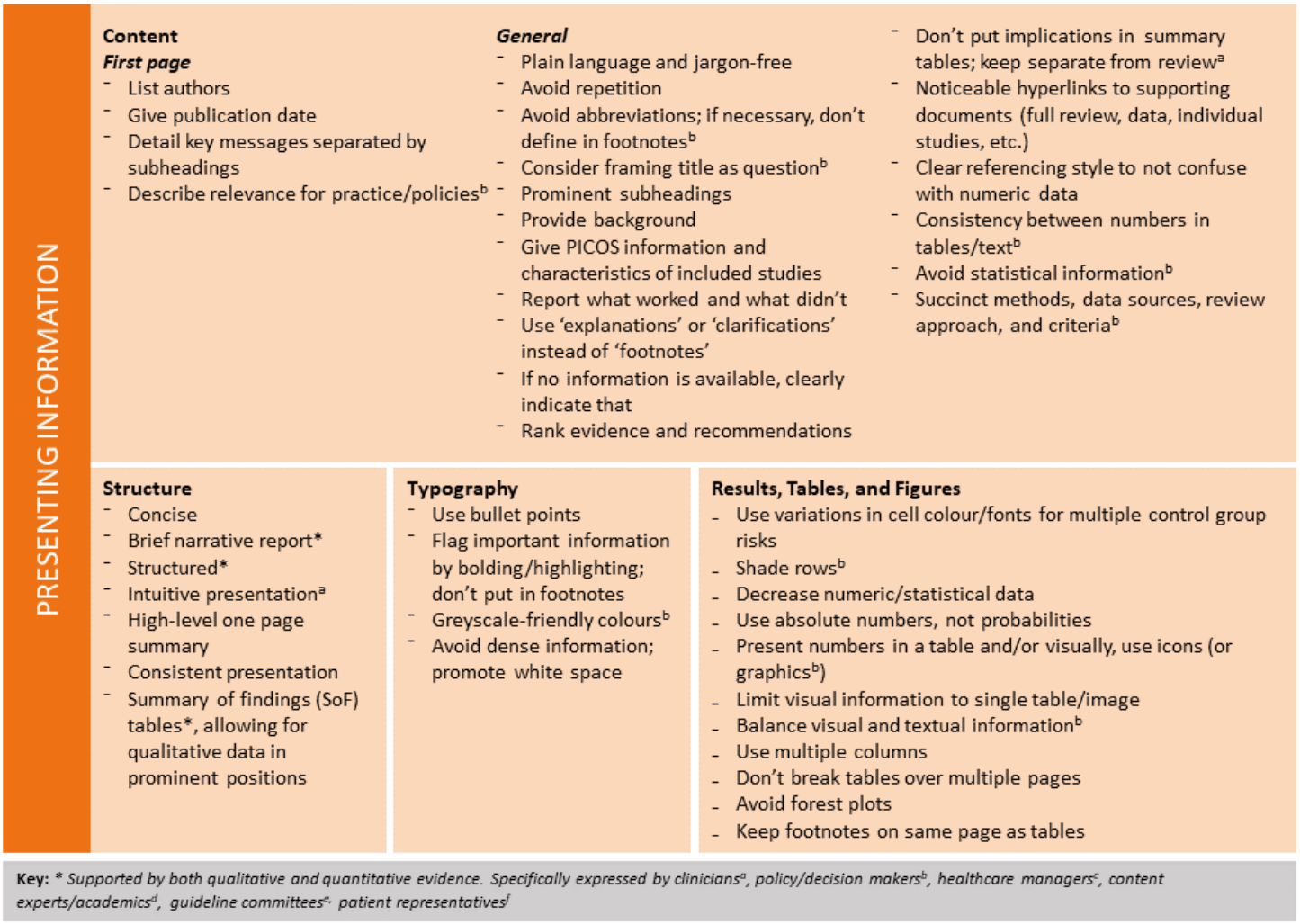


Figure 3

Recommendations for Practice, Presenting Information

Figure 4. Recommendations for Practice, Continued

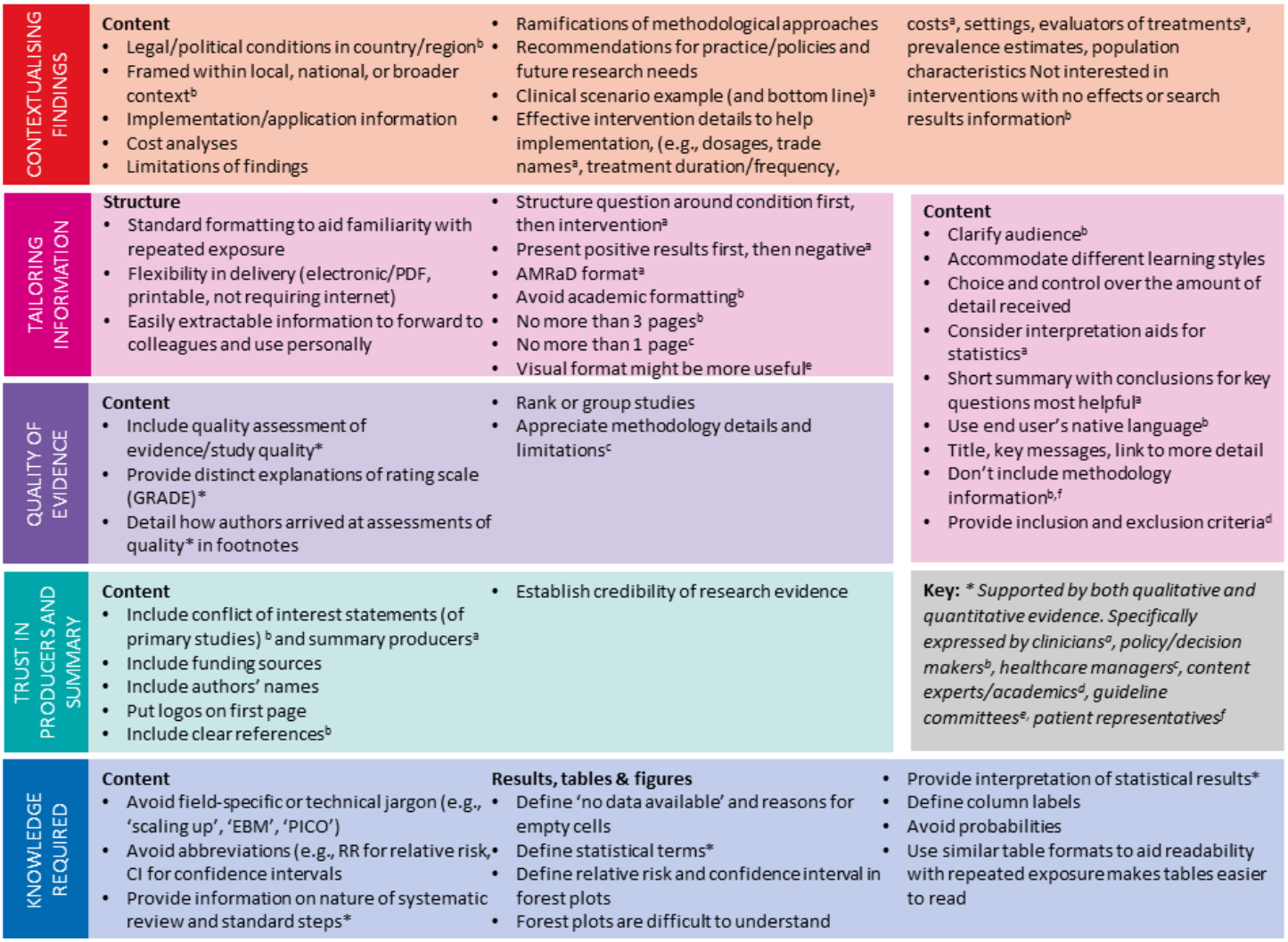


Figure 4

Recommendations for Practice, Continued

Figure 5. Recommendations for Practice, Specialised Reviews

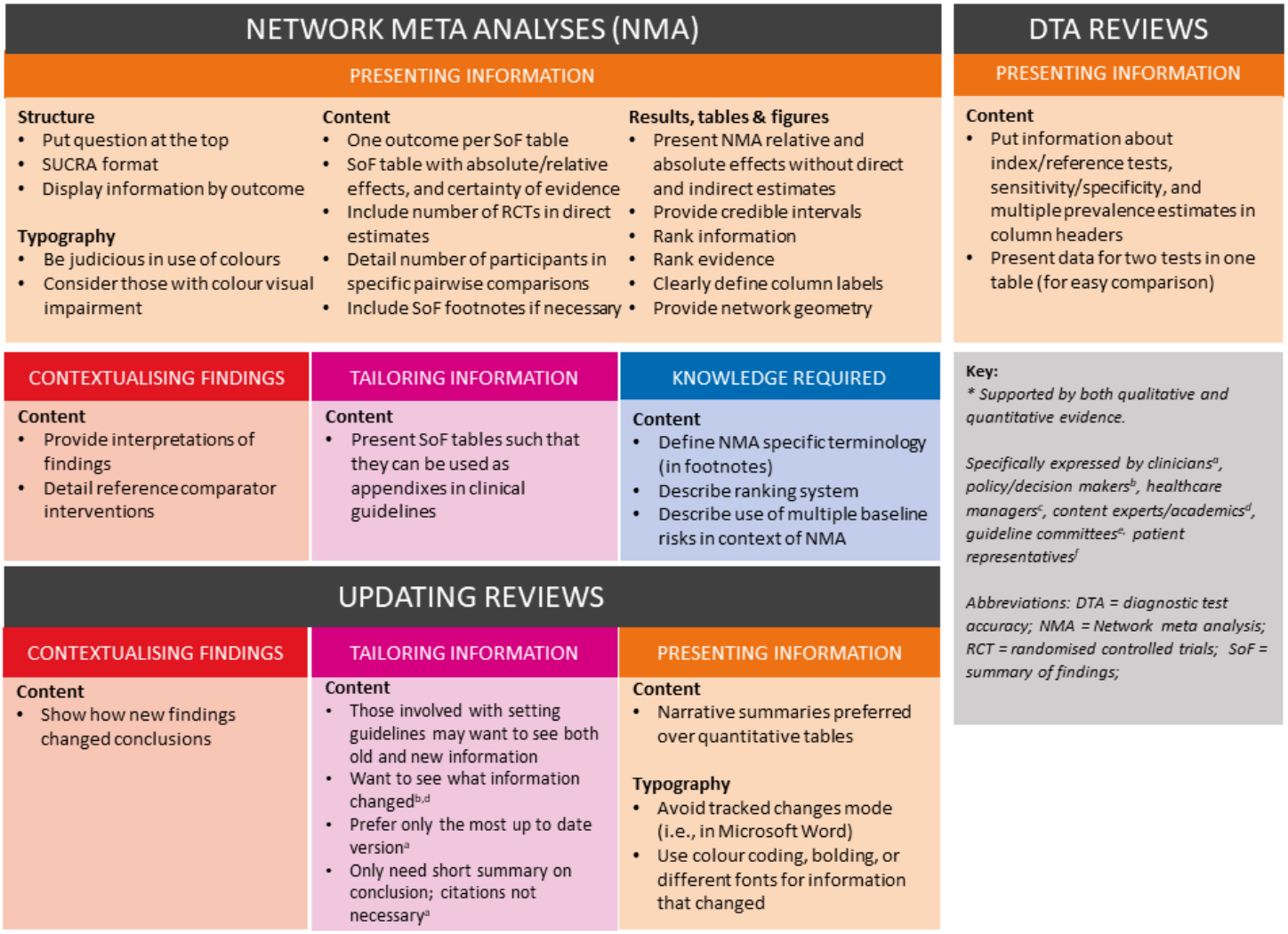


Figure 5

Recommendations for Practice, Specialised Reviews

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendix1PRISMA2020checklist.docx
- Appendix2SearchStrategyOutput.docx
- Appendix3JBICritAppIncludedStudies.xlsx
- Appendix4QuantitativeResults.docx
- Appendix5QualitativeRecommendations.docx