The use of GRADE-CERQual in qualitative evidence synthesis: an evaluation of fidelity and reporting

Wainwright, Megan; Islamiah Zahroh, Rana ; Tunçalp, Özge; Booth, Andrew; Noyes, Jane; Chen, Weilong; Munte-Kaas, Heather; Lewin, Simon

Health Research Policy and Systems

DOI:
10.1186/s12961-023-00999-3

Published: 25/07/2023

Peer reviewed version

The use of GRADE-CERQual in qualitative evidence synthesis: an evaluation of fidelity and reporting

Megan Wainwright[1*], Rana Islamiah Zahroh[2], Özge Tunçalp[3], Andrew Booth[4], Meghan A. Bohren[2], Jane Noyes[5], Weilong Cheng[6], Heather Munthe-Kaas[7], Simon Lewin[8,9,10]

***Corresponding author: Megan Wainwright,** megan@meganwainwright.ca

Affiliations

[1]Department of Anthropology, Faculty of Social Sciences and Health, Durham University, South Road, Durham, United Kingdom.

[2]Gender and Women's Health Unit, Centre for Health Equity, School of Population and Global Health, The University of Melbourne, Carlton, Victoria, Australia

[3]UNDP/UNFPA/UNICEF/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction (HRP), Department of Sexual and Reproductive Health and Research, World Health Organization, Geneva, Switzerland

[4]School of Health and Related Research (ScHARR), Faculty of Medicine, Dentistry and Health, University of Sheffield, Sheffield, United Kingdom

[5]School of Medical and Health Sciences, Bangor University, Bangor, Wales, United Kingdom

[6]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Carlton, Victoria, Australia

[7]The Centre for Epidemic Interventions Research, Norwegian Institute of Public Health, Oslo, Norway

[8]Division of Health Services and Centre for Epidemic Interventions Research (CEIR), Norwegian Institute of Public Health, Oslo, Norway

[9]Health Systems Research Unit, South African Medical Research Council, Cape Town, South Africa

[10] Department of Health Sciences Ålesund, Norwegian University of Science and Technology (NTNU), Norway

ORCID & Email

- Megan Wainwright: 0000-0001-9584-9944; megan@meganwainwright.ca
- Rana Islamiah Zahroh: 0000-0001-7831-2336; r.zahroh@unimelb.edu.au
- Özge Tunçalp: 0000-0002-5370-682X; tuncalpo@who.int
- Andrew Booth: 0000-0003-4808-3880; a.booth@sheffield.ac.uk
- Meghan A. Bohren: 0000-0002-4179-4682; meghan.bohren@unimelb.edu.au
- Jane Noyes: 0000-0003-4238-5984; jane.noyes@bangor.ac.uk
- Weilong Cheng: 0000-0002-5647-4614; weilongc@student.unimelb.edu.au
- Heather Munthe-Kaas: 0000-0002-6051-9010; heather.munthe-kaas@fhi.no
- Simon Lewin: 0000-0001-7521-9515; simonarnold.lewin@fhi.no

Abstract

Background

GRADE-CERQual (Confidence in the Evidence from Reviews of Qualitative Research) is a methodological approach to systematically and transparently assess how much confidence decision makers can place in individual review findings from qualitative evidence syntheses. The number of reviews applying GRADE-CERQual is rapidly expanding in guideline and other decision-making contexts. The objectives of this evaluation were, firstly, to describe the uptake of GRADE-CERQual in qualitative evidence synthesis by review authors and, secondly, to assess both reporting of and fidelity to the approach.

Methods

The evaluation had two parts. Part 1 was a citation analysis and descriptive overview of the literature citing GRADE-CERQual. Authors worked together to code and chart the citations, first by title and abstract and second by full text. Part 2 was an assessment and analysis of fidelity to, and reporting of, the GRADE-CERQual approach in included reviews. We developed fidelity and reporting questions and answers based on the most recent guidance for GRADE-CERQual and then used NVivo12 to document assessments in a spreadsheet and code full-text pdf articles for any concerns that had been identified. Our assessments were exported to Excel and we applied count formulae to explore patterns in the data. We employed a qualitative content analysis approach in NVivo12 to sub-coding all the data illustrating concerns for each reporting and fidelity criteria.

Results

233 studies have applied the GRADE-CERQual approach, with most (n=225, 96.5%) in the field of health research. Many studies (n=97/233, 41.6%) were excluded from full fidelity and reporting assessment because they demonstrated a serious misapplication of GRADE-CERQual, for example interpreting it as a quality appraisal tool for primary studies or reviews. For the remaining studies that applied GRADE-CERQual to assess confidence in review findings, the main areas of reporting concern involved terminology, labelling and completeness. Fidelity concerns were identified in more than half of all studies assessed.

Conclusions

GRADE-CERQual is being used widely within qualitative evidence syntheses and there are common reporting and fidelity issues. Most of these are avoidable and we highlight these as gaps in knowledge and guidance for applying the GRADE-CERQual approach.

Keywords

decision-making, evaluation, fidelity, GRADE-CERQual, methodology, qualitative evidence synthesis, qualitative research, reporting, systematic review

Background

GRADE-CERQual (Confidence in the Evidence from Reviews of Qualitative Research) is a methodological approach for systematically and transparently assessing how much confidence decision makers can place in individual review findings from qualitative evidence syntheses (QES) (also called systematic reviews of qualitative research). The GRADE-CERQual approach first emerged during the development of the OptimizeMNH guidelines (2012)(1), the first World Health Organization (WHO) guideline to formally incorporate QES in its evidence base, and from which the first QES was published in the Cochrane Library (2). The authors of the first QES, together with the GRADE Working Group (https://www.gradeworkinggroup.org), the Cochrane Qualitative and Implementation Methods Group (https://methods.cochrane.org/qi/), and the WHO Department of Reproductive Health and Research, developed the first version of GRADE-CERQual that included only two components – Methodological Limitations and Coherence. The two component GRADE-CERQual was disseminated via a conference presentation (3) and published systematic reviews (4–7). The two component version was very soon replaced by the four component version of GRADE-CERQual which added the relevance and adequacy components to the two component version. It is described in Chapter 21 in the Cochrane handbook (8,9). A fifth potential component (dissemination bias) is currently under development (10–12) but not yet included in the GRADE-CERQual approach.

Over the course of its development, the originators of GRADE-CERQual used a human-centred design approach to further develop the current version, including stakeholder engagement and consensus building sessions in 2014 and 2015, piloting the approach in a number of reviews, gathering feedback via online surveys sent to review teams who had applied the approach, and

developing a global network of collaborators across diverse institutions to support further consensus development. A coordinating group (12 members) and project group (more than 200 members) within the wider GRADE Project Group were established to continue developing and testing the methods for assessing the confidence in qualitative evidence (13).

The first official guidance on the four component GRADE-CERQual approach was published in 2015 in *PLOS Medicine* (9). In 2018, the project group published more detailed guidance to help review authors operationalise GRADE-CERQual in their reviews. The guidance is published as a seven paper supplement in *Implementation Science* that includes a paper for each component and the overall assessment (10,14–19). Beyond the formally published guidance, and adoption by leading global systematic review organisations and guideline producers, dissemination of the GRADE-CERQual approach has included monthly introductory and question and answer webinars, a newsletter, mentorship of review teams, a shared drive of templates to assist users in applying the approach, regular in-person and virtual training workshops, a train the trainer workshop, and translation of the approach into other languages (20–23). Recent years have seen development of a free online data management software to assist review authors with applying the approach - the GRADE-CERQual interactive Summary of Qualitative Findings (iSoQ) tool (24).

In summary, GRADE-CERQual is a method applied in the late stages of a qualitative evidence synthesis, following completion of analysis and synthesis and after review findings have been drafted. Findings are written-up in detail in the body of the manuscript, and shorter summaries of review findings are drafted for inclusion in the GRADE-CERQual Evidence Profile and Summary of Qualitative Findings (SoQF) tables (15), that form two key outputs from the approach. These two tables incorporate synthesis findings, confidence assessments, explanations and contributing studies in a user-friendly format for decision makers. First, reviewers assess each of the four GRADE-CERQual components by identifying any concerns that could reduce their confidence in the review finding. They make a judgement as to whether there are no or very minor, minor, moderate or

serious concerns for each component and record each judgement in the Evidence Profile table, accompanied by an explanation. Once each individual component is assessed, the reviewers proceed to an overall assessment of confidence, considering assessments made for all four components, and decide whether any identified concerns lower their confidence in the review finding. The reviewers start from the assumption that confidence in the finding is high, and then, based on identified concerns for each component, justify downgrading the level of confidence to moderate, low or very low. The guidance papers provide step-by-step instructions for assessing each component and for making the overall assessment (14–19).

The number of reviews applying GRADE-CERQual is expanding rapidly, especially as the approach is now formally recommended for all Cochrane qualitative evidence syntheses and is used widely within WHO and several other global and national guideline-producing agencies including the UK National Institute for Health and Care Excellence (NICE). With the rapid uptake of the approach, members of the GRADE-CERQual coordinating team sought to undertake an evaluation of how the method has been used and, in doing so, to develop fidelity and reporting criteria to assist reviewers when applying GRADE-CERQual. Similar evaluations have been carried out for GRADE (25–33) but this is the first instance for GRADE-CERQual. As qualitative evidence synthesis expands, methodological evaluations are emerging and increasingly needed, including evaluations of reporting standards (34,35), of the use of theories and frameworks (36,37) and specific QES methods or components of methods (38–40). An evaluation of the use of GRADE-CERQual is critical for further improving the application of the approach within reviews conducted for WHO, Cochrane and other global and national agencies.

The objective of this evaluation was twofold. First, to describe the uptake to date of GRADE-CERQual by authors of QES - for example, how often, in what types of reviews, and in which fields of study. Second, to assess how authors report their GRADE-CERQual assessments and to assess fidelity to the most recent guidance for applying the GRADE-CERQual published in 2018 in the journal

Implementation Science (herewith referred to as “the guidance”). We intended to highlight good examples of use, identify common fidelity and reporting concerns, assist review authors to avoid these in the future, and identify where further methodological research, training and guidance is needed.

Methods

This evaluation was conceived as a citation analysis and methods overview. First, it involved identification and descriptive analysis of all citations of GRADE-CERQual. Second, all syntheses that applied the four component version of GRADE-CERQual underwent a detailed examination of their conduct and reporting, and the extent to which fidelity to the published guidance was demonstrated. Methodologically, this evaluation fits within the recent evaluative 'tradition' for assessing reporting guidelines, most notably the PRISMA statement (41). For example, Page and Moher (42), the latter being one of the founders of the reporting guidelines movement, identified over 100 'evaluations of the uptake and impact' of the PRISMA statement and its various extensions. In evaluating 'uptake' they specifically point to the role of citations. We acknowledge however that the GRADE-CERQual approach extends beyond reporting guidance by specifying procedures for conduct of the grading of qualitative findings. Nevertheless, we found citation analysis was the best available approach for evaluating uptake given that GRADE-CERQual can be used in non-published reviews, can be used but not cited, or can be cited but not used.

Our working definition of fidelity is: the extent to which use of GRADE-CERQual demonstrates integrity to the procedures for conduct and reporting as specified in the corpus of GRADE-CERQual methodological papers (10,14–19). While we opted for the term fidelity, Page & Moher(42) preferred the term 'adherence' which they defined as “the extent to which SRs [systematic reviews] comply with each item in the [PRISMA] statement” (p.4). Our choice of fidelity over adherence reflects the fact that while PRISMA carries the status of an almost-mandatory requirement, GRADE-CERQual is largely a value-added procedure.

It was especially important for the team to be reflexive in this evaluation as it was conceptualised and led by members of the GRADE-CERQual coordinating team (43). The team was therefore well qualified to judge fidelity but limited by prior researcher allegiance (44,45) to the approach. This was balanced by: using a coordinating team member (MW) and non-coordinating team members (RIZ, WC) to conduct the fidelity and reporting assessments; ensuring that evaluation team members did not evaluate reviews they co-authored to avoid potential bias (or spin) (44); and using the published guidance papers as the benchmark. Also, we acknowledge that GRADE-CERQual requires ongoing development and improvement and therefore approached the evaluation with a genuine openness to understanding other review authors' interpretation and application of the guidance.

Publication search

Our aim was to identify all publications that had mentioned GRADE-CERQual or referenced a GRADE-CERQual methods paper. First, we searched for "CERQual" and "GRADE-CERQual" in titles and abstracts in MEDLINE, EMBASE, Scopus, Web of Science, CINAHL and in full-texts in Google Scholar. Search strategies and results are available in Additional File 1. All searches were conducted in August 2020. For papers not using either abbreviation we searched using the following combination of terms: confidence AND evidence AND reviews AND "Qualitative research" (in recognition that "in" and "from" are stop words in most databases). We then conducted forward citation searches for 18 methodological papers (see Additional File 2), including official guidance and descriptions of the GRADE-CERQual method, and known translations, in Google Scholar (using the Publish or Perish tool), Scopus and Web of Science. We deliberated the value of forward citation searches for known published reviews that used GRADE-CERQual but decided that these references would likely duplicate those retrieved by the full-text phrase searches on Google Scholar. No language, date, or geographical filters were applied. Where possible, MEDLINE duplicates were excluded at the database search stage.

Part 1 – Citation Analysis and Descriptive Overview

Stage 1.1: Coding and Charting: titles and abstracts

We imported details of authors, year, title, journal, volume, issues, pages, DOI, abstract and URL for all retrieved papers from Endnote into Excel and divided the citations between the co-authors. Co-authors independently coded their assigned references in Excel for language, type of publication and whether a GRADE-CERQual coordinating team member was a co-author (see Additional File 3 for coding questions and answers). Twenty percent of articles assigned to types of publication other than "review/synthesis" were independently co-coded by a second researcher to assess whether unlabelled syntheses/reviews had eluded full-text coding/charting. As a team, we decided to exclude theses and dissertations to focus on comparable units of reporting detail from within the peer-reviewed literature. Spanish, Portuguese, French, Korean, Mandarin, Swedish, and Norwegian publications were coded by authors proficient in each language. We used Google Translate to translate abstracts and titles in other languages.

Stage 1.2: Coding and Charting: full texts

We uploaded publications coded to "Review/Synthesis (quantitative or qualitative)" at the title/abstract stage to the Covidence online screening platform (46). We prepared an extraction form for full-text coding including guidance cited, field of study, whether GRADE-CERQual was applied, and type of review (Additional File 3). Each publication was coded by one author. Where the team had previously misattributed a publication as a review or synthesis when it was a protocol, a dissertation, a methodological, or conceptual paper, etc., then it was returned to the title/abstract coding stage. All such publications were then reviewed by the first author who verified the misattribution and updated the title/abstract coding and charting spreadsheet. We were lenient when interpreting "Was GRADE-CERQual applied to review findings?"; we selected "yes" if authors simply stated that they had used or applied GRADE-CERQual, even when the application was unclear. Co-authors proficient in that language coded all publications in languages other than English

and 20% of English language publications were checked by a second author. MW and RIZ, who had coded most of the publications, also checked all references coded by other authors to further improve consistency.

Part 2: Fidelity and Reporting Assessment and Analysis

Stage 2.1: Coding for serious fidelity issues

When planning the evaluation, we anticipated that we would flag reviews that applied GRADE-CERQual to quantitative systematic reviews as having serious fidelity issues and that these would therefore not proceed to full reporting and fidelity assessment. However, close reading of each study revealed a need to flag other serious deficiencies ("fatal flaws") in fidelity to the approach. Our definitive list of fatal flaws is shown in Table 1. Only studies that applied GRADE-CERQual to assess confidence in individual review findings and were supported exclusively by findings from qualitative research proceeded to full reporting and fidelity assessment. At this stage we also coded for review type, synthesis method, and co-authorship or acknowledgement of members of the GRADE-CERQual coordinating team. In addition, we flagged studies thought to demonstrate an innovative use of the GRADE-CERQual approach.

<Insert Table 1. Questions and responses for identifying fatal flaws>

Stage 2.2: Reporting and Fidelity Assessment

We developed and finalised the reporting and fidelity assessment criteria (Additional File 4) through multiple rounds of pilot testing. First, we developed an expanded version of previously published minimum criteria for fidelity to the GRADE-CERQual approach (14). All co-authors then piloted this version on two publications. Second, we discussed our experiences and feedback, and decided to separate assessment into 'reporting' and 'fidelity'. Third, the first author redrafted the criteria and these were then piloted by some co-authors on a third article. Finally, we compared and discussed our assessments to confirm further modifications.

MW and RIZ built an NVivo12 (47) master project in which the reporting and fidelity assessment criteria would be applied. Each publication was added to both an overall case node ("evaluated studies") and to a case node specific to the publication. We created a case classification sheet to register responses to each reporting and fidelity question. We also created a node for each criterion and data extracts were coded to these nodes to justify concerns with fidelity or reporting. We then created a framework matrix in NVivo12 to summarise concerns regarding fidelity. MW and RZ conducted simultaneous real-time pilot-testing, followed by minor revisions of the NVivo12 project structure against a single paper. MW and RIZ then independently pilot-tested the framework on two publications and met to compare assessments. Very minor revisions sought to improve the clarity of the responses to each question and a further two papers were independently assessed. MW and RIZ subsequently divided the remaining studies between them for single author assessment.

For analysis, RIZ looked for patterns in the data by exporting the case classification sheet to Excel and running counts of the data. MW applied a conventional qualitative content analysis approach (48) to sub-coding all the data illustrating concerns for each reporting and fidelity criteria. Qualitative content analysis was considered appropriate for descriptively and inductively identifying categories of concerns for each reporting and fidelity criteria and for observing their prevalence. In presenting our results we have deliberately avoided 'naming and shaming' individual studies. Instead, we seek to offer constructive feedback, to help review authors improve their reporting and fidelity and to identify how the GRADE-CERQual project group can further improve guidance for review authors.

Results

A total of 2732 publications were found from database searches. After duplicates were removed, 1312 publications were assessed at the title/abstract stage. We identified 461 publications as reviews or evidence synthesis at the full text stage, of which 233 (50.5%) publications applied the four component version of GRADE-CERQual. Out of these 233 publications, a total of 136

publications (58.4%) applied GRADE-CERQual to individual review findings and were assessed for fidelity and reporting. Figure 1 shows the number of publications identified and included across the different stages of the evaluation. Overall, the number of published studies that applied GRADE-CERQual to individual review findings is increasing (Figure 2).

<Insert Figure 1. PRISMA flowchart.>

<Insert Figure 2. Number of published studies that applied GRADE-CERQual to individual review findings by year.>

Part 1. Citation Analysis and Descriptive Overview

There were 462 records coded as a review or synthesis (either quantitative, qualitative or mixed-methods) and proceeded to full-text coding and charting. Quantitative reviews were included at this stage in recognition that GRADE-CERQual is occasionally wrongly applied to findings from quantitative effect reviews, and because we wanted to ensure that we captured this issue in the evaluation. Of the 461 reviews or syntheses, 251 (54.4%) actually applied GRADE-CERQual in their reviews, of which 233 publications (92.8% of those applying the approach) used the four component version, and 18 (7.2%) publications used the earlier two component version. All results of the title and abstract coding are available in Additional File 5. As reviews that used the four component version of GRADE-CERQual are the most relevant to this evaluation, we only present results for these 233 studies.

Most of these 233 studies used standard review or synthesis methods for synthesising primary qualitative research (n=223, 95.7%), while seven (3%) used overview or umbrella review approaches and three (1.2%) used scoping review approaches. Currently there is no official guidance for applying GRADE-CERQual to overview or umbrella reviews or scoping reviews. We included these in our evaluation as an opportunity to learn how review authors adapted the guidance. Although 184 studies (78.9%) were published from 2018 onwards, the *PLOS Medicine* methods paper (9) remained

the most commonly-cited guidance (n=120, 51.5%). Twenty-six publications (11.1%) neither cited methodological guidance publications (9,10,14,15,15–19), nor cited a published review or other publication about GRADE-CERQual. Almost all (n=223, 95.7%) of the publications were written in the English language, with the remainder in Mandarin (n=5, 2.1%), Spanish (n=3, 1.2%), Norwegian (n=1, 0.4%), and Swedish (n=1, 0.4%). See Additional File 6 for full results of full-text coding and charting.

Part 2. Fidelity and Reporting Assessment and Analysis

Serious Misapplications of GRADE-CERQual

A total of 233 studies applied the four component version of GRADE-CERQual. We categorised 97 (41.6%) as having seriously misused GRADE-CERQual. Fatal flaws, to borrow a term used in critical appraisal, include fundamentally misinterpreting GRADE-CERQual as either: a quality appraisal tool of included studies (n=19, 19.6%), a critical appraisal tool or reporting tool for the review itself (n=11, 11.3%), or an approach to assess general confidence in all findings or in the review as a whole (rather than individual review findings) (n=13, 13.4%). A very serious fatal flaw was that 11 studies (11.3%) wrongly applied GRADE-CERQual to findings from a quantitative systematic review. The last and most common fatal flaw (n=43, 44.3%) was applying GRADE-CERQual to review findings developed from <u>both</u> qualitative and quantitative data. GRADE-CERQual is designed to be applied to review findings supported exclusively by qualitative data (that is data emanating from qualitative data collection methods and qualitative methods of analysis). Therefore, application of GRADE-CERQual to review findings supported by both qualitative and quantitative data (including descriptive quantitative studies like surveys), does not align with current guidance. A mixed-methods study or mixed-methods review can apply GRADE-CERQual to the findings of a qualitative evidence synthesis that has been conducted in parallel to a quantitative review. Guidance from the GRADE-CERQual project group does not yet exist for applying GRADE-CERQual to review findings supported by mixed evidence within convergent synthesis designs.. Most of the 43 studies with this fatal flaw neither acknowledged that this application deviateed from the guidance nor did they sufficiently

explain how they adapted the assessment of each component to the quantitative origins of the data. It was challenging to identify that review findings were supported by qualitative and quantitative data with reporting gaps proving common. For example, many did not clearly label that they were mixed-methods or mixed studies reviews and did not specify the design used for their synthesis (e.g., convergent, parallel). Although the current version of GRADE-CERQual has not been designed for use in a scoping review or an overview/umbrella review we included those types of reviews in this evaluation as an opportunity to learn how review authors had adapted the guidance. However, all three scoping reviews and six of the eight overview/umbrella reviews were excluded for one of the above mentioned fatal flaws. As was to be expected, the two umbrella reviews that went on to full assessment scored poorly on fidelity to the guidance.

Most Common Reporting Issues

A total of 136 studies applied the four-component version of GRADE-CERQual to assess confidence in individual review findings and were consequently brought to the next stage of analysis. Full results of the reporting assessment of these 136 studies are available in Additional File 7. We synthesised the reporting issues into three broad areas: labelling, terminology, and completeness. Figure 3 summarises the main reporting concerns. The most common reporting issues relate to two key outputs; the Summary of Qualitative Findings (SoQF) Table and the Evidence Profile. Though most reviews included one or other of these they were often poorly labelled and omitted crucial elements such as an explanation for the overall assessment and the citations for the studies contributing to each summarised review finding.

<Insert Figure 3. Main reporting concerns grouped by broad area>

Most common fidelity issues

As seen on Figure 4a fidelity concerns were identified in over half of all studies assessed. Publications citing the 2018 *Implementation Science* guidance (14–19) were less likely to generate

concerns regarding each fidelity criteria compared to publications citing the 2015 guidance (9) (Figure 4b). It is important to note that in this evaluation we assessed all studies against the 2018 guidance. The largest improvements were seen in the greater clarity with which the GRADE-CERQual approach and each of the four components were conceptualised. This was to be expected given that the 2018 series published four separate papers detailing how to operationalise each component (see section "Steps" in each of these publications (16–19)). Mixed methods reviews tended to occasion more concerns, on average, than qualitative evidence syntheses.

<Insert Figure 4. Graph showing concerns per fidelity criteria#

#Figure 4a. Bar graph showing number of studies with concerns on each fidelity question; Figure 4b. Line graph comparing percentage of concerns between studies that cited 2015 and 2018 guidance for the GRADE-CERQual approach.>

Although the frequencies on Figure 4 help to visualise patterns, they do encompass very minor through to serious concerns. For example, any difference from the guidance was flagged as a concern, whether very minor or not. To address this, the qualitative content analysis sub-coding revealed the full range of concerns for each criteria (full results in Additional File 8). Table 2 is an abbreviated version highlighting the most common, alongside what we considered to be the most important, concerns. We found that these identified concerns often reflected wider conceptual or operational concerns in the review as a whole. In the following sections we discuss and reflect on these cross-cutting issues.

<Insert Table 2. Common concerns identified on fidelity assessment>

Of the 136 studies we assessed for fidelity, the most common synthesis methods used were thematic synthesis (n=69, 50.7%), followed by meta-ethnography (n=18, 13.2%). The coordinating team has long acknowledged that experience to date has focused on applying GRADE-CERQual to descriptive findings with a need for more application to interpretive or analytic findings. It is

encouraging to see many meta-ethnographies emerging that apply GRADE-CERQual. However, fidelity varied widely between meta-ethnographies, demonstrating a need for clearer guidance on applying GRADE-CERQual to these types of syntheses. While we did not identify any examples with a complete absence of concerns, four meta-ethnographies had no concerns for most criteria (49–52).

Examples of Good Fidelity

We did not identify any concerns for eleven publications. Six of these were Cochrane Reviews (53–58) and two were reports of commissioned reviews (59,60) where review authors are given sufficient word limits to fully describe the GRADE-CERQual approach and demonstrate how it was applied. In comparison, journal articles may have more restrictive word limits. All of the Cochrane reviews involved GRADE-CERQual Coordinating Team Members as either co-authors or acknowledged mentors. The remaining four publications were published in academic journals (61–64) and all included additional files. As writing within word limits is a serious challenge with publishing systematic reviews, including details of GRADE-CERQual assessments (such as Evidence Profile or SoQF tables) is crucial for demonstrating fidelity to the approach and transparency in analysis. Since fidelity in conceptualising and operationalising the assessment of each of the four components needs to be improved, we recommend the following six studies as good examples. Collectively, these studies included evidence profiles for their GRADE-CERQual assessments and did not raise any concerns for how they conceptualised the components (52,65–69).

Discussion

Our evaluation identified several broad issues which can better inform users of the GRADE-CERQual approach, including guideline developers, review authors and researchers. These issues include conceptual challenges, problems with the application of the approach, and inadequate reporting that compromises transparency. The following discussion considers each in turn.

Conceptual challenges related to applying GRADE-CERQual

Confusion between confidence in review findings and quality

GRADE-CERQual is an approach for assessing confidence in, rather than the quality of, review findings. In our evaluation, we found several indicators of conceptual confusion between 'quality' and 'confidence'. These included the words being used interchangeably; review authors describing the GRADE-CERQual approach in the section addressing quality appraisal of primary studies; and suggestions that GRADE-CERQual was intended to assess the quality of review findings. However, in the context of qualitative evidence syntheses, the "quality" of a review finding is not a useful concept - it is more appropriate to consider how much confidence we have in a review finding. Further, while assessing the methodological limitations of studies included in a synthesis (sometimes called 'quality appraisal') is a key stage of the review process, this should not be confused with assessing confidence in a review finding. A review finding is typically based on data from several studies, each with their own methodological limitations. As we have described elsewhere (16), the methodological limitations component of GRADE-CERQual requires that review authors have previously applied a critical appraisal tool to all contributing studies, and then consider the importance of these methodological limitations in relation to each review finding.

Poor understanding of the kind of review GRADE-CERQual can and cannot be applied to

Eleven reviews of quantitative evidence misapplied the GRADE-CERQual approach; typically based on the misinterpretation that GRADE-CERQual can be applied to narrative summaries of quantitative results (i.e. a synthesis without meta-analysis). A synthesis without meta-analysis should not be described as 'qualitative' as this contributes to confusion, and in fact recent guidance for reporting quantitative synthesis without meta-analysis deliberately avoids the term narrative synthesis altogether (70). The bottom line is GRADE-CERQual has been designed for assessment of findings grounded in data from primary studies that used qualitative methods for collecting and analysing

data. GRADE for effectiveness is the appropriate approach for review results based on quantitative data, even in cases where a meta-analysis was not possible (71). Numerous approaches are emerging to guide users in selecting an analysis approach appropriate to the aim and underlying data of their systematic review (70,72). Furthermore, considering that the most common fatal flaw was applying GRADE-CERQual to review findings developed from both qualitative and quantitative data (e.g. survey data), points to a gap in available methods for assessing confidence in findings from mixed methods or mixed studies reviews using convergent designs (where quantitative data is transformed into qualitative form and analysed and synthesised together with qualitative data).

Problems with the application of the approach

Specific review findings versus overall themes

GRADE-CERQual is applied to individual review findings. In some reviews, these findings may be organised into broader theme headings but, in these cases, GRADE-CERQual should still be applied to the granular review findings. For example, “inequality in access to information” could be a theme identified in a synthesis, and this theme may include specific review findings - for instance, “Unlike women under 40, women over 40 faced technological barriers to accessing the information they needed on their diagnosis”. In this case GRADE-CERQual should be applied to the specific review finding(s) and not the theme category. However, the authors of 14 reviews appeared to apply GRADE-CERQual to overall themes and subthemes, instead of the summaries of specific review findings. Instead of presenting a summary of a review finding in the SoQF or Evidence Profile, only a brief theme label (no more than 3-5 words) was provided. This is not congruent with the guidance because the summary of the finding needs to provide sufficient detail to be understood by readers of the tables. One way to demonstrate in an SoQF or Evidence Profile that some review findings fall within the same theme is to divide these tables into sections corresponding to these theme headings (73), or to include the brief theme label alongside a summary of the relevant findings (74).

Quantifying GRADE-CERQual assessments

The quantification of GRADE-CERQual assessments proved to be one of the most concerning fidelity issues. In the most extreme form of quantification, evidence profile tables only contained numbers of studies without explanations. For example, a review might state how many studies achieved each quality rating for methodological limitations, and how many indirectly or partially relevant studies contributed to relevance. While reference to the quantity of studies may help to explain an assessment, that assessment itself is not a count. The assessment for each component is a judgement of appropriate concerns regarding that component. Similarly, the overall assessment of confidence as either high, moderate, low or very low is a judgement, not a count. It was concerning that some authors determined confidence to be "high" based on whether the finding was supported by a given number of studies or cases. This deviation from the guidance completely overlooks the need to factor-in concerns noted for all four components when making the overall assessment. Some authors also deviated from the guidance by establishing scoring systems to determine overall confidence – e.g., low confidence for one component with serious concerns, or two components with moderate concerns. Scoring was considered a concern rather than an innovation on the GRADE-CERQual approach because of over-reliance on quantification and the false sense of precision that this provides (15).

Assessing components at the individual study rather than the review finding level

The error of quantifying assessments is also tied closely to a conceptual issue around assessing the GRADE-CERQual components at the individual study level rather than the review finding level. GRADE-CERQual is an assessment of confidence in the review finding. Thus, while component assessment involves looking back at the included studies supporting the finding, the review finding is being assessed, not the contributing studies. The categories of no or very minor, minor, moderate, or serious concerns are applied to the review finding, not to individual supporting studies. Some review authors applied the components to the individual study, assigning each contributing study a level of concern for each component, and then reporting a count of how many supporting studies

registered each level of concern. The focus is deflected towards the sheer number of concerns aggregated from primary studies, and away from a careful consideration of the importance that review authors assign to these concerns in relation to the specific review finding that they are assessing.

GRADE-CERQual components not assessed in terms of concerns

A further issue relates to authors not conceptualising component assessments in terms of concerns at all, as was the case in 36 reviews. Some authors confused the assessment categories for the overall assessment of confidence with the assessment categories for component assessments. For example, instead of indicating no or very minor concerns about relevance, authors might assign high or moderate relevance. This departs from the guidance given, "our aim is not to judge whether data [component] has been achieved, but to judge whether there are grounds for concern regarding [the component] that are serious enough to lower our confidence in the review finding" (18). In other instances, authors just used the component name as the assessment – e.g., "adequate data" rather than "no/very minor concerns for adequacy of data". This again points to a subtle but nonetheless significant tendency to frame component assessments in terms of whether an ideal has been met, rather than identifying grounds for concern.

Issues specific to each component

The evaluation identified several issues specific to each GRADE-CERQual component (methodological limitations, coherence, relevance, and adequacy of data). The most common fidelity concern related to coherence being wrongly defined as "consistency within and across studies". Other frequent concerns related to reporting of methodological limitations and how such limitations were considered in relation to each review finding. Considerations of relevance and adequacy focused on a limited interpretation of these components (e.g., for adequacy focusing on either quantity or richness but not both, and for relevance focusing only on one aspect like setting (thus overlooking other aspects of context like perspective, population or phenomenon of interest). Some

review authors utilised concepts related to one component to assess another. These component-specific issues are described in detail in Additional File 9.

Reporting that compromises transparency

Inadequate reporting compromises transparency. GRADE-CERQual assessments are judgements made by review authors, therefore, transparency is key to understanding these judgements and is therefore fundamental to the overall approach. To demonstrate adherence to the principle of transparency, review authors must provide an explanation for each of their component assessments and their overall assessment of confidence. Summarising assessments for each component within an explanation for the overall assessment cannot replace individual explanations for each component. Furthermore, an overall assessment is incomplete without an explanation. We recommend that, as a minimum, the explanation for the overall assessment should state the level of concern for each component. But we encourage authors to add additional detail about the concerns that are driving down their confidence in the finding. An important part of transparency is including SoQF table and evidence profiles. Assessing fidelity based only on a SoQF table is very difficult given insufficient detail for the reader to understand how review authors interpreted and assessed each component. We therefore recommend that qualitative evidence syntheses include Evidence Profiles as additional files.

Strengths and Limitations

Here we reflect on the strengths and limitations, firstly, of the evaluation, and, secondly, of the GRADE-CERQual approach. A strength of this evaluation is that we systematically assessed each included review against specific criteria, and have clearly identified fatal flaws and most common reporting and fidelity issues, with the aim of being as useful as possible to future review authors. We consider this evaluation an essential complement to our published methodological guidance. Limitations of our evaluation include being reliant only on published materials (we did not have the

resources to contact review authors for additional information), and having potentially missed relevant applications of GRADE-CERQual in our pragmatic decision to exclude theses and dissertations. Possible issues related to applying the approach in the context of a degree-related output were not identified.

This evaluation shows that an important strength of the GRADE-CERQual approach is that review authors have found GRADE-CERQual to be applicable to a wide range of topics and to different types of qualitative evidence synthesis. Limitations of the approach include that it has not yet been adapted for mixed methods reviews with convergent designs. The current version of GRADE-CERQual is not intended to be applied to such review findings but there is no alternative approach for review authors to use. This is an important area for the future development of the GRADE-CERQual approach. Furthermore, review authors could benefit from summaries of the published guidance, focusing in particular on the "steps" for assessment section of the papers. The new GRADE-CERQual iSoQ tool is expected to assist review authors with following the steps in the approach.

Conclusions – Future Agendas

This study identified the most common and serious reporting and fidelity concerns when applying the GRADE-CERQual approach. The fidelity and reporting criteria used for the evaluation can help review authors to avoid common pitfalls and improve alignment with the guidance. This evaluation has also identified key areas for future research (Table 3), and future priorities for the dissemination and implementation of the GRADE-CERQual approach (Table 4). Future research could evaluate uptake and use beyond review authors, to include for example the use of GRADE-CERQual assessments by decision makers.

List of Abbreviations

- GRADE-CERQual: Confidence in the Evidence from Reviews of Qualitative research

- iSoQ: GRADE-CERQual interactive Summary of Qualitative Findings (iSoQ) tool
- NICE: National Institute for Health and Care Excellence
- QES: Qualitative evidence synthesis
- SoQF Table: Summary of Qualitative Findings Table

List of Additional Information

- Additional file 1.docx – Search log: GRADE-CERQual topic and citation searches
- Additional file 2.docx – List of GRADE-GRADE-CERQual publications for citation searching
- Additional file 3.docx – Coding and charting questions and answers
- Additional file 4.docx – Reporting and fidelity assessment criteria
- Additional file 5.docx – Results of title and abstract coding and charting
- Additional file 6.docx – Results of full-text coding and charting that applied GRADE-CERQual
- Additional file 7.docx – GRADE-CERQual reporting assessment results
- Additional file 8.docx – GRADE-CERQual reporting and fidelity assessment sub-coding of identified concerns by criteria
- Additional file 9.docx – Issues specific to each component

Declarations

Ethics approval and consent to participate

The data for this study consisted of published systematic reviews and therefore ethics approval was not needed.

Consent for publication

There was no direct individual contact involved in this study, therefore consent for publication is not needed.

Availability of data and materials

Additional information files are provided, and further data may be provided upon reasonable request.

Competing interests

Most of the evaluation authors (MW, OT, SL, AB, MAB, JN, HMK) are Coordinators of the GRADE-CERQual Project Group coordinating team and have been involved closely in developing and promoting the approach. In addition, some of the evaluation authors were also authors of some of the included reviews. As evaluation authors, we were keen to understand how the GRADE-CERQual approach has been applied to date, including both concerns with and innovations in its use. However, we acknowledge both the limitations and strengths of our interests in the approach. The authors declare that they have no other competing interests.

Funding

This study was commissioned and paid for by the UNDP/UNFPA/UNICEF/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction (HRP), a cosponsored program executed by the World Health Organization (WHO). Copyright in the original work on which this article is based belongs to WHO. The authors have been given permission to publish this article. The authors alone are responsible for the views expressed in this article and they do not necessarily represent the views, decisions, or policies of the World Health Organization. SL receives additional funding from the South African Medical Research Council. MAB is supported by an Australian Research Council Discovery Early Career Researcher Award (DE200100264) and a Dame Kate Campbell Fellowship (University of Melbourne Faculty of Medicine, Dentistry, and Health Sciences).

Authors' contributions

The study was conceptualised by members of the GRADE-CERQual coordinating team (MW, SL, OT, JN, AB, MAB, HMK). MW designed the study and methodology with input from co-authors. MW, RIZ, SL, OT, MAB, JN, AB and HMK conducted study screening and extraction. MW and RIZ undertook the fidelity and reporting assessments. WC joined the author team to screen and assess the studies published in Mandarin with the support of RIZ. MW and RIZ conducted the analysis. MW drafted the manuscript for publication and co-authors contributed to the content and revision of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. WHO | Optimizing health worker roles for maternal and newborn health [Internet]. WHO. World Health Organization; [cited 2022 Jan 21]. Available from: https://www.who.int/reproductivehealth/publications/maternal_perinatal_health/9789245043/en/

2. Gülmezoglu AM, Chandler J, Shepperd S, Pantoja T. Reviews of qualitative evidence: a new milestone for Cochrane. Cochrane Database of Systematic Reviews [Internet]. 2013 [cited 2022 Jan 21];(11). Available from: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.ED000073/full?highlightAbstract=editorial%7C*chandler%7Ceditori

3. Lewin S, Glenton C, Munthe-Kaas H, Carlsen B, Colvin CJ, Rashidian A. Assessing how much certainty to place in findings from qualitative evidence syntheses: the GRADE-CERQual approach. In: Better Knowledge for Better Health | Un meilleur savoir pour une meilleure santé. Québec City, Canada: John Wiley & Sons; 2013.

4. Glenton C, Colvin C, Carlsen B, Swartz A, Lewin S, Noyes J, et al. Barriers and facilitators to the implementation of lay health worker programmes to improve access to maternal and child health: qualitative evidence synthesis. Cochrane Database of Systematic Reviews [Internet]. 2013 [cited 2022 Jan 21];(2). Available from: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD010414/full

5. Colvin CJ, de Heer J, Winterton L, Mellenkamp M, Glenton C, Noyes J, et al. A systematic review of qualitative evidence on barriers and facilitators to the implementation of task-shifting in midwifery services. Midwifery. 2013 Oct 1;29(10):1211–21.

6. Jordan J, Rose L, Dainty KN, Noyes J, Blackwood B. Factors that impact on the use of mechanical ventilation weaning protocols in critically ill adults and children: a qualitative

evidence-synthesis. Cochrane Database of Systematic Reviews [Internet]. 2016 [cited 2022 Jan 21];(10). Available from: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD011812.pub2/full

7. Munthe-Kaas H, Hammerstrøm K, Nordlund KR. Effekt av og erfaringer med kontinuitetsfremmende tiltak i barnevernsinstitusjoner [Internet]. Norwegian Institute of Public Health; [cited 2022 Jan 21]. Available from: https://www.fhi.no/en/publ/2013/effects-and-experiences-of-interventions-to-promote-continuity-in-residenti/

8. Chapter 21: Qualitative evidence [Internet]. [cited 2022 Jan 21]. Available from: https://training.cochrane.org/handbook/current/chapter-21

9. Lewin S, Glenton C, Munthe-Kaas H, Carlsen B, Colvin CJ, Gülmezoglu M, et al. Using Qualitative Evidence in Decision Making for Health and Social Interventions: An Approach to Assess Confidence in Findings from Qualitative Evidence Syntheses (GRADE-CERQual). PLOS Medicine. 2015 Oct 27;12(10):e1001895.

10. Booth A, Lewin S, Glenton C, Munthe-Kaas H, Toews I, Noyes J, et al. Applying GRADE-CERQual to qualitative evidence synthesis findings–paper 7: understanding the potential impacts of dissemination bias. Implement Sci. 2018 Jan 25;13(1):12.

11. Toews I, Glenton C, Lewin S, Berg RC, Noyes J, Booth A, et al. Extent, Awareness and Perception of Dissemination Bias in Qualitative Research: An Explorative Survey. PLOS ONE. 2016 Aug 3;11(8):e0159290.

12. Toews I, Booth A, Berg RC, Lewin S, Glenton C, Munthe-Kaas HM, et al. Further exploration of dissemination bias in qualitative research required to facilitate assessment within qualitative evidence syntheses. Journal of Clinical Epidemiology. 2017 Aug 1;88:133–9.

13. Munthe-Kaas HM. Appendix 10.1 Methods used to develop the GRADE-CERQual approach: Developing the GRADE-CERQual approach: Report from stakeholder brainstorming sessions. Exploring the development of the GRADE-CERQual and TRANSFER approaches to enhance accountability for reasonableness in evidence-informed decision-making processes. [Norway]: University of Oslo. Institute of Health and Science, Faculty of Medicine; 2021.

14. Lewin S, Booth A, Glenton C, Munthe-Kaas H, Rashidian A, Wainwright M, et al. Applying GRADE-CERQual to qualitative evidence synthesis findings: introduction to the series. Implement Sci. 2018 Jan 25;13(1):2.

15. Lewin S, Bohren M, Rashidian A, Munthe-Kaas H, Glenton C, Colvin CJ, et al. Applying GRADE-CERQual to qualitative evidence synthesis findings—paper 2: how to make an overall CERQual assessment of confidence and create a Summary of Qualitative Findings table. Implement Sci. 2018 Jan 25;13(1):10.

16. Munthe-Kaas H, Bohren MA, Glenton C, Lewin S, Noyes J, Tunçalp Ö, et al. Applying GRADE-CERQual to qualitative evidence synthesis findings—paper 3: how to assess methodological limitations. Implement Sci. 2018 Jan 25;13(1):9.

17. Colvin CJ, Garside R, Wainwright M, Munthe-Kaas H, Glenton C, Bohren MA, et al. Applying GRADE-CERQual to qualitative evidence synthesis findings—paper 4: how to assess coherence. Implement Sci. 2018 Jan 25;13(1):13.

18. Glenton C, Carlsen B, Lewin S, Munthe-Kaas H, Colvin CJ, Tunçalp Ö, et al. Applying GRADE-CERQual to qualitative evidence synthesis findings—paper 5: how to assess adequacy of data. Implement Sci. 2018 Jan 25;13(1):14.

19. Noyes J, Booth A, Lewin S, Carlsen B, Glenton C, Colvin CJ, et al. Applying GRADE-CERQual to qualitative evidence synthesis findings–paper 6: how to assess relevance of the data. Implement Sci. 2018 Jan 25;13(1):4.

20. Organización Mundial de la Salud. Manual para la elaboración de directrices– 2a edición: 14. Recomendaciones fuertes cuando las pruebas científicas son de baja calidad [Internet]. Organización Mundial de la Salud; 2015 [cited 2022 Jan 21]. Available from: https://apps.who.int/iris/handle/10665/333776

21. Lewin S, Glenton C, Munthe-Kaas HM, Carlsen B, Colvin CJ, Gulmezoglu AM, et al. Uso de Evidência Qualitativa em Decisões para Intervenções Sociais e em Saúde: Uma Abordagem para Avaliar a Confiança em Achados de Sínteses de Evidências Qualitativas (GRADE-CERQual). PLoS Med [Internet]. 2015;12(10). Available from: https://aagts.brasilia.fiocruz.br/wp-content/uploads/2020/03/GRADE-GRADE-CERQual_PLOS_portugu%C3%AAs.pdf

22. World Health Organization. WHO handbook for guideline development: 15. Usando evidências de pesquisa qualitativa para desenvolver diretrizes da OMS [Internet]. Geneva: World Health Organization; 2018 [cited 2022 Jan 21]. Available from: https://www.who.int/publications/guidelines/WHO_Handbook-for-Guideline-Develoment-Chapter-15_Portuguese.pdf?ua=1

23. Toma TS, Barreto JOM. GRADE-CERQual: uma abordagem para avaliar a confiança nos resultados de sínteses de evidências qualitativas. GRADE-CERQual: uma abordagem para avaliar a confiança nos resultados de sínteses de evidências qualitativas. 2019;220 p-220 p.

24. GRADE-CERQual Interactive Summary of Qualitative Findings (iSoQ) [Internet]. Oslo, Norway: Norwegian Institute of Public Health (developed by the Epistemonikos Foundation, Megan Wainwright Consulting and the Norwegian Institute of Public Health for the GRADE-CERQual Project Group); Available from: isoq.epistemonikos.org

25. Agoritsas T, Merglen A, Heen AF, Kristiansen A, Neumann I, Brito JP, et al. UpToDate adherence to GRADE criteria for strong recommendations: an analytical survey. BMJ Open. 2017 Nov 1;7(11):e018593.

26. Wyer PC. From MARS to MAGIC: The remarkable journey through time and space of the Grading of Recommendations Assessment, Development and Evaluation initiative. Journal of Evaluation in Clinical Practice. 2018;24(5):1191–202.

27. Mercuri M, Baigrie BS. What confidence should we have in GRADE? Journal of Evaluation in Clinical Practice. 2018;24(5):1240–6.

28. Mercuri M, Gafni A. The evolution of GRADE (part 1): Is there a theoretical and/or empirical basis for the GRADE framework? Journal of Evaluation in Clinical Practice. 2018;24(5):1203–10.

29. Mercuri M, Gafni A. The evolution of GRADE (part 2): Still searching for a theoretical and/or empirical basis for the GRADE framework. Journal of Evaluation in Clinical Practice. 2018;24(5):1211–22.

30. Mercuri M, Gafni A. The evolution of GRADE (part 3): A framework built on science or faith? Journal of Evaluation in Clinical Practice. 2018;24(5):1223–31.

31. Mercuri M, Baigrie B, Upshur REG. Going from evidence to recommendations: Can GRADE get us there? Journal of Evaluation in Clinical Practice. 2018;24(5):1232–9.

32. Miles KE, Rodriguez R, Gross AE, Kalil AC. Strength of Recommendation and Quality of Evidence for Recommendations in Current Infectious Diseases Society of America Guidelines. Open Forum Infectious Diseases. 2021 Feb 1;8(2):ofab033.

33. Weiss MJ, Hornby L, Shemie SD, Appleby A, Rochwerg B. GRADEing the un-GRADE-able: a description of challenges in applying GRADE methods to the ethical and implementation questions of pediatric organ donation guidelines. Journal of Clinical Epidemiology. 2018 Nov 1;103:134–7.

34. France EF, Ring N, Thomas R, Noyes J, Maxwell M, Jepson R. A methodological systematic review of what's wrong with meta-ethnography reporting. BMC Medical Research Methodology. 2014 Nov 19;14(1):119.

35. Lawson DO, Leenus A, Mbuagbaw L. Mapping the nomenclature, methodology, and reporting of studies that review methods: a pilot methodological review. Pilot and Feasibility Studies. 2020 Jan 30;6(1):13.

36. Bergström A, Ehrenberg A, Eldh AC, Graham ID, Gustafsson K, Harvey G, et al. The use of the PARIHS framework in implementation research and practice—a citation analysis of the literature. Implementation Science. 2020 Aug 27;15(1):68.

37. Field B, Booth A, Ilott I, Gerrish K. Using the Knowledge to Action Framework in practice: a citation analysis and systematic review. Implementation Science. 2014 Nov 23;9(1):172.

38. France EF, Uny I, Ring N, Turley RL, Maxwell M, Duncan EAS, et al. A methodological systematic review of meta-ethnography conduct to articulate the complex analytical phases. BMC Medical Research Methodology. 2019 Feb 18;19(1):35.

39. Munn Z, Dias M, Tufanaru C, Porritt K, Stern C, Jordan Z, et al. The 'quality' of JBI qualitative research synthesis: a methodological investigation into the adherence of meta-aggregative systematic reviews to reporting standards and methodological guidance. JBI evidence synthesis. 2021 Apr 8;19(5):1119–39.

40. Tsang A, Maden M. CLUSTER searching approach to inform evidence syntheses: A methodological review. Research Synthesis Methods. 2021 Jun 1;12:576–89.

41. 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. BMJ. 2021 Mar 29;372:n71.

42. Page MJ, Moher D. Evaluations of the uptake and impact of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) Statement and extensions: a scoping review. Systematic Reviews. 2017 Dec 19;6(1):263.

43. Olmos-Vega FM, Stalmeijer RE, Varpio L, Kahlke R. A practical guide to reflexivity in qualitative research: AMEE Guide No. 149. Medical Teacher. 2023 Mar 4;45(3):241–51.

44. Lieb K, Osten-Sacken J von der, Stoffers-Winterling J, Reiss N, Barth J. Conflicts of interest and spin in reviews of psychological therapies: a systematic review. BMJ Open. 2016 Apr 1;6(4):e010606.

45. Uttley L. 56 Understanding bias, conflicts of interest and researcher allegiance in systematic reviews. BMJ Evidence-Based Medicine. 2019 Jul 1;24(Suppl 1):A33–4.

46. Covidence systematic review software [Internet]. Melbourne, Australia: Veritas Health Innovation; 2021 [cited 2021 Oct 21]. Available from: https://www.covidence.org/

47. QSR International Pty Ltd. NVivo (Version 12) [Internet]. 2018. Available from: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home

48. Hsieh HF, Shannon SE. Three Approaches to Qualitative Content Analysis. Qual Health Res. 2005 Nov 1;15(9):1277–88.

49. Fernández-Basanta S, Movilla-Fernández MJ, Coronado C, Llorente-García H, Bondas T. Involuntary Pregnancy Loss and Nursing Care: A Meta-Ethnography. International Journal of Environmental Research and Public Health. 2020 Jan;17(5):1486.

50. Fernández-Basanta S, Coronado C, Movilla-Fernández MJ. Multicultural coping experiences of parents following perinatal loss: A meta-ethnographic synthesis. Journal of Advanced Nursing. 2020;76(1):9–21.

51. Kingdon C, Downe S, Betran AP. Interventions targeted at health professionals to reduce unnecessary caesarean sections: a qualitative evidence synthesis. BMJ Open. 2018 Dec 1;8(12):e025073.

52. Wolstenholme H, Kelly C, Hennessy M, Heary C. Childhood fussy/picky eating behaviours: a systematic review and synthesis of qualitative studies. International Journal of Behavioral Nutrition and Physical Activity. 2020 Jan 3;17(1):2.

53. Ames HM, Glenton C, Lewin S. Parents' and informal caregivers' views and experiences of communication about routine childhood vaccination: a synthesis of qualitative evidence. Cochrane Database Syst Rev. 2017 Feb 7;2:CD011787.

54. Ames HM, Glenton C, Lewin S, Tamrat T, Akama E, Leon N. Clients' perceptions and experiences of targeted digital communication accessible via mobile devices for reproductive, maternal, newborn, child, and adolescent health: a qualitative evidence synthesis. Cochrane Database of Systematic Reviews [Internet]. 2019 [cited 2022 Jan 21];(10). Available from: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD013447/full

55. Bohren MA, Berger BO, Munthe-Kaas H, Tunçalp Ö. Perceptions and experiences of labour companionship: a qualitative evidence synthesis. Cochrane Database of Systematic Reviews [Internet]. 2019 [cited 2022 Jan 21];(3). Available from: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD012449.pub2/full

56. Munabi-Babigumira S, Glenton C, Lewin S, Fretheim A, Nabudere H. Factors that influence the provision of intrapartum and postnatal care by skilled birth attendants in low- and middle-income countries: a qualitative evidence synthesis. Cochrane Database of Systematic Reviews [Internet]. 2017 [cited 2022 Jan 21];(11). Available from: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD011558.pub2/full

57. Odendaal WA, Watkins JA, Leon N, Goudge J, Griffiths F, Tomlinson M, et al. Health workers' perceptions and experiences of using mHealth technologies to deliver primary healthcare services: a qualitative evidence synthesis. Cochrane Database of Systematic Reviews [Internet]. 2020 [cited 2022 Jan 21];(3). Available from: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD011942.pub2/full

58. Houghton C, Meskell P, Delaney H, Smalle M, Glenton C, Booth A, et al. Barriers and facilitators to healthcare workers' adherence with infection prevention and control (IPC) guidelines for respiratory infectious diseases: a rapid qualitative evidence synthesis. Cochrane Database of Systematic Reviews [Internet]. 2020 [cited 2022 Jan 21];(4). Available from: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD013582/full

59. Sopory P, Novak J. Quarantine as a Non-Pharmaceutical Intervention: Qualitative Research Evidence Synthesis [Internet]. National Academies of Sciences, Engineering, and Medicine; 2020 p. 27. Available from: https://www.nationalacademies.org/documents/embed/link/LF2255DA3DD1C41C0A42D3BEF0989ACAECE3053A6A9B/file/D5B43CC7112B9EF109EED0ABDD22A6D72BEA6D18B943

60. Sopory P, Novak J. Public Health Emergency Operations Coordination: Qualitative Research Evidence Synthesis [Internet]. National Academies of Sciences, Engineering, and Medicine; 2020 p. 27. Available from: https://www.nationalacademies.org/documents/embed/link/LF2255DA3DD1C41C0A42D3BEF0989ACAECE3053A6A9B/file/D5B43CC7112B9EF109EED0ABDD22A6D72BEA6D18B943

61. Bohren MA, Vogel JP, Tunçalp Ö, Fawole B, Titiloye MA, Olutayo AO, et al. Mistreatment of women during childbirth in Abuja, Nigeria: a qualitative study on perceptions and experiences of women and healthcare providers. Reprod Health [Internet]. 2017 Jan 17 [cited 2021 Apr 6];14. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5240205/

62. Bulthuis SE, Kok MC, Raven J, Dieleman MA. Factors influencing the scale-up of public health interventions in low- and middle-income countries: a qualitative systematic literature review. Health Policy Plan. 2020 Mar 1;35(2):219–34.

63. Carstensen K, Lou S, Groth Jensen L, Konstantin Nissen N, Ortenblad L, Pfau M, et al. Psychiatric service users' experiences of emergency departments: a CERQual review of qualitative studies. Nord J Psychiatry. 2017 May;71(4):315–23.

64. Herbert AC, Ramirez AM, Lee G, North SJ, Askari MS, West RL, et al. Puberty Experiences of Low-Income Girls in the United States: A Systematic Review of Qualitative Literature From 2000 to 2014. J Adolesc Health. 2017 Apr;60(4):363–79.

65. Diaw M, Sibeoni J, Manolios E, Gouacide JM, Brami C, Verneuil L, et al. The Lived Experience of Work-Related Issues Among Oncology Nurses: A Metasynthesis. Cancer Nurs. 2020 Jun;43(3):200–21.

66. Egerton T, Diamond LE, Buchbinder R, Bennell KL, Slade SC. A systematic review and evidence synthesis of qualitative studies to identify primary care clinicians' barriers and enablers to the management of osteoarthritis. Osteoarthritis Cartilage. 2017 May;25(5):625–38.

67. Esteban E, Coenen M, Ito E, Gruber S, Scaratti C, Leonardi M, et al. Views and Experiences of Persons with Chronic Diseases about Strategies that Aim to Integrate and Re-Integrate Them

into Work: A Systematic Review of Qualitative Studies. Int J Environ Res Public Health. 2018 May 18;15(5):E1022.

68. Hughes NM, Noyes J, Eckley L, Pritchard T. What do patients and family-caregivers value from hospice care? A systematic mixed studies review. BMC Palliative Care. 2019 Feb 8;18(1):18.

69. Jones V, Crowe M. How people from ethnic minorities describe their experiences of managing type-2 diabetes mellitus: A qualitative meta-synthesis. Int J Nurs Stud. 2017 Nov;76:78–91.

70. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ. 2020 Jan 16;368:l6890.

71. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. Journal of Clinical Epidemiology. 2011 Apr 1;64(4):383–94.

72. Noyes J, Booth A, Moore G, Flemming K, Tunçalp Ö, Shakibazadeh E. Synthesising quantitative and qualitative evidence to inform guidelines on complex interventions: clarifying the purposes, designs and outlining some methods. BMJ Global Health. 2019 Jan 1;4(Suppl 1):e000893.

73. Hennelly N, Cooney A, Houghton C, O'Shea E. Personhood and Dementia Care: A Qualitative Evidence Synthesis of the Perspectives of People With Dementia. Gerontologist. 2021 Apr 3;61(3):e85–100.

74. Chamberlain C, Ralph N, Hokke S, Clark Y, Gee G, Stansfield C, et al. Healing The Past By Nurturing The Future: A qualitative systematic review and meta-synthesis of pregnancy, birth and early postpartum experiences and views of parents with a history of childhood maltreatment. PLoS One. 2019;14(12):e0225441.

List of Tables

Table 1. Questions and responses for identifying fatal flaws.

Questions	Responses
How was GRADE-CERQual used?	**CORRECT** 1. Applied to assess confidence in individual review findings
	FATAL FLAWS 2. Applied as an assessment of general confidence in review findings, or the review itself as a whole 3. Applied as a critical appraisal tool for included studies 4. Applied as a critical appraisal or reporting tool for the review
What type of data were underlying the review findings that GRADE-CERQual was applied to?	**CORRECT** 1. Qualitative
	FATAL FLAWS 2. Qualitative and Quantitative 3. Quantitative only

842

843 ***Table 2. Common concerns identified on fidelity assessment.***

Fidelity Assessment Questions **Common concerns identified in bullet points**
1. The authors demonstrate an accurate conceptualisation of GRADE-CERQual (that is, an approach for assessing confidence in the findings of a qualitative evidence synthesis). • Appears under quality appraisal section • Sometimes referred to as a tool to assess quality of findings or evidence • Conceptualised as an assessment of contributing studies
2.The authors have made an overall assessment of confidence based on the assessment of all four components • No mention of the 4 components at all • Some components not assessed • Applied own scoring rules for determining level of assessment
3.The authors applied GRADE-CERQual to individual review findings • Applied GRADE-CERQual at the study level not finding level • Applied it to short theme or category titles
4.Authors conceptualise methodological limitations in line with the guidance • Applied the levels of concern to individual studies rather than review findings • Conceptualised the assessment as a count of appraisal categories, not specific limitations in relation to the finding • Component not defined and no Evidence Profile or SoQF tables from which to infer • Not conceptualised in terms of identifying concerns • Problems with how critical appraisals were done (e.g., only yes or no, no explanation) • Specific methodological limitations mentioned but not how important they are in relation to the finding
5. Authors conceptualise coherence in line with the guidance • Component not defined and no Evidence Profile or SoQF tables from which to infer • No demonstration of thinking of it in terms of the fit between review finding and data from primary studies, only focus on primary studies • Not conceptualised in terms of identifying concerns • Using wrong definition (“Consistent within and across studies”) • Assessment was quantified
6. Authors conceptualise adequacy of data in line with the guidance • Component not defined and no Evidence Profile or SoQF table from which to infer • Not assessed in terms of concerns • Not assessing both quantity and richness, emphasising one or the other • Confounding with other components • Quantify the assessment of the component
7. Authors conceptualise relevance in line with the guidance • Component not defined and no Evidence Profile or SoQF tables from which to infer • Language of concerns not used, or not used correctly • Not all elements of ‘context’ were considered in the assessment • Quantify the assessment by counting how many primary studies are indirect or partial, not using concerns
8.The GRADE-CERQual assessments are presented in-line with the guidance for SoQF tables and or Evidence Profiles • No SoQF or Evidence Profile tables included

• Key elements missing or left out (such as references or explanations) • Way of writing explanations for component or overall assessments not aligned with guidance
9. Summarised review findings were produced in line with the guidance • Summaries of findings either too detailed or too brief • Just theme or category names, not summarised review findings

844

845 ***Table 3. Future research agenda for the GRADE-CERQual approach.***

• Refine the criteria that need to be met for authors to claim fidelity to the GRADE-CERQual approach, drawing on the assessment criteria used in this evaluation • Develop reporting criteria for the use of GRADE-CERQual in qualitative evidence syntheses, drawing on the reporting assessment criteria used in this evaluation • Clarify guidance on the types of analytic outputs, or findings, from a qualitative evidence synthesis to which GRADE-CERQual should be applied, including in the context of a meta-ethnography. • Improve the guidance on applying each GRADE-CERQual component, focusing on: (1) helping authors to conceptually navigate the boundary between considering concerns at the individual study level and assessing the importance of these concerns at the review finding level; (2) navigating the conceptual boundaries between components • Develop more detailed guidance on how to bring together GRADE-CERQual component assessments to make an overall assessment of confidence, while avoiding quantification of individual component assessments • Evaluate the impact of the new GRADE-CERQual interactive Summary of Qualitative Findings (iSoQ) tool on the completeness and consistency of reporting of GRADE-CERQual assessments, including explanations of these, and on the presentation of Summary of Qualitative Findings and Evidence Profile tables • Undertake methodological research on how to apply GRADE-CERQual in the context of mixed methods syntheses (including convergent designs) in order to develop guidance on this

846

847 ***Table 4. Future dissemination and implementation agenda for the GRADE-CERQual approach.***

• Improve the accessibility of guidance on component assessments and making an overall assessment, for example by disseminating concise instructions for these • Encourage review authors, and journals publishing qualitative evidence syntheses, to include SoQF tables in all qualitative evidence syntheses and to make GRADE-CERQual Evidence Profiles available as supplementary materials • Explore how good examples identified in this evaluation of the application of GRADE-CERQual can be used to support training and other forms of capacity strengthening for applying the approach • Continue to identify examples of GRADE-CERQual applications that adhere to our current reporting and fidelity criteria so that these can be used for capacity strengthening • Explore how training and guidance materials linked to the new GRADE-CERQual interactive Summary of Qualitative Findings (iSoQ) tool could be used to strengthen capacity to apply the GRADE-CERQual approach appropriately • Encourage the incorporation of training and guidance materials on the GRADE-CERQual approach into teaching modules on qualitative evidence synthesis and on systematic reviewing more broadly

848