

Optimising the value of the critical appraisal skills programme (CASP) tool for quality appraisal in qualitative evidence synthesis

Research Methods in Medicine &
Health Sciences
2020, Vol. 1(1) 31–42
© The Author(s) 2020



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2632084320947559
journals.sagepub.com/home/rmm



Hannah A Long , David P French and Joanna M Brooks

Abstract

The value of qualitative evidence synthesis for informing healthcare policy and practice within evidence-based medicine is increasingly recognised. However, there is a lack of consensus regarding how to judge the methodological quality of qualitative studies being synthesised and debates around the extent to which such assessment is possible and appropriate. The Critical Appraisal Skills Programme (CASP) tool is the most commonly used tool for quality appraisal in health-related qualitative evidence syntheses, with endorsement from the Cochrane Qualitative and Implementation Methods Group. The tool is recommended for novice qualitative researchers, but there is little existing guidance on its application. This article considers issues related to the suitability and usability of the CASP tool for quality appraisal in qualitative evidence synthesis in order to support and improve future appraisal exercises framed by the tool. We reflect on our practical experience of using the tool in a systematic review and qualitative evidence synthesis. We discuss why it is worth considering a study's underlying theoretical, ontological and epistemological framework and how this could be incorporated into the tool by way of a novel question. We consider how particular features of the tool may impact its interpretation, the appraisal results and the subsequent synthesis. We discuss how to use quality appraisal results to inform the next stages of evidence synthesis and present a novel approach to organising the synthesis, whereby studies deemed to be of higher quality contribute relatively more to the synthesis. We propose tool modifications, user guidance, and areas for future methodological research.

Keywords

Qualitative evidence synthesis, systematic review, critical appraisal, quality appraisal, Critical Appraisal Skills Programme

Background

Since the 1990s, the systematic review has been considered to be the cornerstone methodology for informing healthcare policy and practice within the evidence-based medicine movement. Systematic review methods were originally developed to identify, appraise and synthesise quantitative evidence from randomised controlled trials in order to answer questions of effectiveness (e.g. “what works?”).¹ Today, the methodology is used to review evidence from studies with different designs, including qualitative, and to answer different types of questions (e.g. “how and why does this work?”).² This shift reflects a broader growing recognition of the unique and useful contribution that qualitative evidence can make to evidence-based medicine.^{2,3}

Methods for the synthesis of primary qualitative research predate formalised systematic reviews of quantitative research.⁴ Nevertheless, the process of including qualitative research in systematic reviews has typically involved adopting most of the methodological stages of systematic reviews of quantitative research (until the data synthesis stage). However, the unquestioning use of systematic review methods

Manchester Centre for Health Psychology, School of Health Sciences, University of Manchester, Manchester, UK

Corresponding author:

Hannah A Long, Manchester Centre for Health Psychology, School of Health Sciences, University of Manchester, Manchester M13 9PL, UK.
Email: hannah.long@manchester.ac.uk

appropriate for traditional quantitative research in the synthesis of primary qualitative research has created certain epistemological and practical challenges,⁵ and some see the two methodologies to be somewhat incompatible.⁶ Booth has proposed an inclusive narrative which recognises a 'dual heritage' of systematic review and qualitative research methods, "whereby some appropriate qualitative evidence synthesis techniques derive from the systematic review of effectiveness, whereas others originate from primary qualitative research"⁷ (p. 20).

A key stage common to all systematic reviews is quality appraisal of the evidence to be synthesized.^{1,8} There is broad debate and little consensus among the academic community over what constitutes 'quality' in qualitative research. 'Qualitative' is an umbrella term that encompasses a diverse range of methods, which makes it difficult to have a 'one size fits all' definition of quality and approach to assessing quality.⁹ There are debates on how to judge the rigour and validity of qualitative evidence (including which quality criteria, concepts or dimensions should be used, if any); issues in treating the broad spectrum of qualitative research as a unified field; and how to appraise qualitative research undertaken from different ontologies and epistemologies, among others.^{8,10–12} However, qualitative researchers are increasingly expected to demonstrate rigour in their research (e.g. for national exercises, such as Excellence in Research in Australia,¹³ the Evaluation of Research Quality in Italy,¹⁴ and the Research Excellence Framework (REF) in the UK,¹⁵ which aim to assess and rate the quality of research undertaken at higher education institutions, to inform research funding allocations and university standings at a national and international level). Thus, there is need to develop appropriate standards by which to judge quality in order to avoid the application of inappropriate criteria.

An element of the broader debate is whether quality appraisal of studies included in qualitative evidence synthesis is possible, appropriate and defensible, and whether to use quality judgements to inform the next stages of synthesis and if so how. In a systematic review of 82 qualitative evidence syntheses in the field of healthcare, published between 2005 and 2008, Hannes and Macaitis found that appraisal results were most commonly used to exclude studies deemed to be of a lower quality.¹⁶ Other uses include sensitivity analysis to assess the impact of including lower quality studies on the synthesis findings¹⁷ and weighting the synthesis in favour of findings from studies determined to be of a higher quality.¹⁸

With these debates in mind, we wish to take a pragmatic approach towards the issue of quality appraisal in qualitative evidence synthesis. Increased support for

and inclusion of qualitative evidence in evidence-based medicine and healthcare policy and practice is cause for celebration by the qualitative community. If qualitative evidence is to have a seat at the decision-making table of healthcare policy and practice, policymakers and practitioners must be able to have confidence in the quality of qualitative research. It therefore follows that some form of quality appraisal in qualitative evidence synthesis is appropriate and it would appear that there is some degree of agreement on this front.

The Cochrane Qualitative and Implementation Methods Group (hereon Cochrane) have published guidance for conducting qualitative and mixed-method evidence synthesis.^{2,8,19–22} The guidance is dedicated in part to methods for assessing methodological strengths and limitations of qualitative research included in qualitative evidence synthesis.⁸ In this guidance paper, Cochrane explicitly advocate performing quality appraisal and using such assessments to make judgements about the impact of methodological limitations on the synthesised findings. In their systematic review, Hannes and Macaitis found that quality appraisal was performed in the majority (72%) of evidence syntheses.¹⁶ Published five years later, Dalton et al.'s descriptive overview of 145 health and social care-related qualitative evidence syntheses identified quality appraisal exercises in 133 (92%) syntheses.²³ Moreover, many quality concepts and principles have been proposed^{12,24–28} and more than one-hundred criteria- and checklist-based qualitative quality appraisal tools have been identified.²⁹ Cochrane have encouraged further methodological development of quality appraisal through reflection and formal testing.⁸

In response, the present article is a practical and opportunistic reflection on issues encountered when performing quality appraisal as part of a systematic review and qualitative evidence synthesis. Our aim is to discuss the suitability and usability of the Critical Appraisal Skills Programme (CASP) qualitative checklist tool for quality appraisal in qualitative evidence synthesis in order to support and improve future appraisal exercises framed by the tool.³⁰ The CASP tool is the most commonly used checklist/criteria-based tool for quality appraisal in health and social care-related qualitative evidence syntheses^{16,23} and is recommended for novice qualitative researchers,³¹ but there is little existing guidance on applying the tool. We discuss why it is worth considering a qualitative study's underlying theoretical, ontological and epistemological framework and how this could be incorporated into the CASP tool in the form of a novel question. We consider why qualitative knowledge and expertise is needed to appraise qualitative research well, without which particular features of the CASP tool may be applied in a way that is problematic for the appraisal results. We

consider how to use quality appraisal results to inform the next stages of synthesis, by organising the synthesis to prioritise findings from studies deemed to be of a higher quality. In the process, we suggest solutions to potential challenges for researchers using the tool. We propose CASP tool modifications, user guidance and areas for further methodological research.

Method

The context of our quality appraisal

We have previously conducted a systematic review and qualitative evidence synthesis of women's experiences of receiving a false positive test result from breast screening.³² We systematically identified, appraised and synthesised data from eight papers (reflecting seven qualitative primary studies). All primary study authors had used semi-structured interview methods to collect qualitative data on women's personal and subjective feelings, views and beliefs about their breast screening experience and the breast screening service they had attended.

We used the CASP tool to appraise the quality of studies included in our review. The first author (a PhD student with a background in health psychology and some experience of systematic review and qualitative research methods) independently appraised the quality of all included papers. The second and third author (experienced researchers with expertise in health psychology, systematic reviews and qualitative research methods) independently appraised the quality of one paper each. We met to discuss our preliminary quality appraisal decisions, including the issues we had encountered with the tool. Following discussion, we decided to modify the tool in an attempt to address the issues identified. We had not originally planned to do this, but deemed it necessary in order to manage the issues that had surfaced. The first author re-appraised all papers and the third author re-appraised one paper to trial the tool modifications. The three of us met again to discuss these decisions. The first author re-appraised quality for a third time to consolidate the final quality appraisal decisions and justifications for these. We did not exclude any studies from our review based on quality.

The Critical Appraisal Skills Programme tool

The CASP tool is a generic tool for appraising the strengths and limitations of any qualitative research methodology.³⁰ The tool has ten questions that each focus on a different methodological aspect of a qualitative study (Box 1). The questions posed by the tool ask the researcher to consider whether the research

Box 1. The 10 questions of the CASP qualitative checklist tool³⁰

1. Was there a clear statement of the aims of the research?
2. Is a qualitative methodology appropriate?
3. Was the research design appropriate to address the aims of the research?
4. Was the recruitment strategy appropriate to the aims of the research?
5. Was the data collected in a way that addressed the research issue?
6. Has the relationship between researcher and participants been adequately considered?
7. Have ethical issues been taken into consideration?
8. Was the data analysis sufficiently rigorous?
9. Is there a clear statement of findings?
10. How valuable is the research?

methods were appropriate and whether the findings are well-presented and meaningful. The CASP tool was "designed to be used as an educational, pedagogic tool, as part of a workshop setting".³⁰

The CASP tool is considered to be a user-friendly option for a novice qualitative researcher and is endorsed by Cochrane and the World Health Organisation for use in qualitative evidence synthesis.^{8,16,31} We chose the tool partly for these reasons; the first author had no prior experience of formally appraising the quality of qualitative research. Further, the CASP tool was devised for use with health-related research and was therefore deemed appropriate for the context of our review.

The CASP tool has been found to be a relatively good measure of transparency of research practice and reporting standards, but a relatively less good measure of research design and conduct. When compared with different appraisal methods (i.e. a quality framework and unprompted expert judgement), the CASP tool was found to give a good indication of the procedural aspects of a study and the details that should be reported, but to produce lower agreement within and between reviewers compared to the other appraisal methods.³³ Hannes et al. compared the CASP tool with the evaluation tool for qualitative studies (ETQS) and the Joanna Briggs Institute (JBI) tool on five dimensions of validity. The CASP tool was found to be the least sensitive of the three to interpretative, evaluative and theoretical validity and did not compare well as a measure of intrinsic methodological quality.³⁴ However, it has been acknowledged that assessments of study quality are made based on the details included in the study report. It follows that quality appraisal is contingent on adequate reporting, and may only

assess reporting, rather than study conduct.¹⁷ Nevertheless, the CASP tool is the most commonly used checklist/criteria-based tool for quality appraisal in health and social care-related qualitative evidence synthesis.^{16,23} Authors' reasons for using a particular quality appraisal tool are, if reported, often reported only briefly or somewhat generically, and thus the reasons for the CASP tool's prevalence are not known.

Data synthesis

We organised our synthesis based on the results of our quality appraisal. To synthesise our data we followed Thomas and Harden's approach to thematic synthesis,³⁵ modified for the purpose of organising the synthesis by quality. The data to be synthesised were the primary study authors' interpretations of their study findings and all direct quotations from participants. The first author extracted all text from the 'results' or 'findings', and any findings reported in the abstract, and treated this text as data. Thomas and Harden's approach to thematic synthesis involves three main stages: line-by-line coding, the development of descriptive themes and the generation of analytical themes. Our approach to inquiry was informed by dialectical pluralism, which allows for research undertaken from various and competing paradigms (and underpinned by different epistemological and ontological assumptions) to be synthesised into a new, agreed upon whole.³⁶

Results

Appraising a study's approach to inquiry

In an early attempt to appraise the quality of primary studies, we identified issues that could not be sufficiently captured by any of the tool's original questions. Specifically, we encountered issues related to the clarity with which the guiding qualitative paradigms were described by primary study authors and the congruence of paradigms with the study methods. In the majority of cases, the authors' approach to inquiry had not been clearly reported. In some cases, the authors borrowed data analysis techniques associated with another method and applied these to their chosen method without discussion of these decisions. There was evidence of conceptual confusion within the reports, e.g. authors reported that they allowed the data to speak for itself, but chose a data analysis method that is based on the premise that reality is interpreted in various ways, text as data has multiple meanings and thus understanding depends on subjective interpretations. A further example was of authors reporting a mixed approach to inquiry, adopting multiple, diverse and altogether incompatible paradigms (i.e.

phenomenology, grounded theory and discourse analysis) without explanation or justification.

In its original form, the CASP tool does not have a criterion for appraising the clarity and appropriateness of a study's qualitative paradigm. The implications of this have been noted by previous authors.^{29,32} Qualitative approaches to research are informed by diverse theory, epistemology and ontology, and there are multiple philosophical positions that researchers could take. It is considered to be good practice for qualitative researchers to reflect on and clearly report their approach to inquiry as it has implications for how their methodology and methods are understood.³⁷ As a result, others can understand and appropriately assess the quality of the research, as the nature of qualitative research necessitates alternative criteria to more positivist quality concepts often applied in other types of research, such as reliability.³⁸ It therefore follows that consideration of this should be reflected in the framework used to appraise quality.

We deemed it important to our review to include appraisal of the primary study authors' approach to inquiry. We generated a question of our own design and modified the CASP tool (Box 2): are the study's theoretical underpinnings (e.g. ontological and epistemological assumptions; guiding theoretical framework (s)) clear, consistent and conceptually coherent? We developed three hints to accompany the question:

1. To what extent is the paradigm that guides the research project congruent with the methods and methodology, and the way these have been described? (See also Yardley's concept of 'coherence',¹²)
2. To what extent is there evidence of problematic assumptions about the chosen method of data analysis? e.g. assuming techniques or concepts from another method (e.g. use of data saturation, originating in grounded theory) apply to chosen method (e.g. Braun and Clarke's reflexive thematic analysis^{39,40}) without discussion or justification.
3. To what extent is there evidence of conceptual clashes or confusion in the paper? e.g. claiming a constructionist approach but then treating participants' accounts as a transparent reporting of their experience and behaviour.

Interpretation of the tool's response options and 'hints'

We discussed the suitability and usability of the tool's design for quality appraisal. The tool's questions are designed to be prompts for assessing generic features of qualitative research. The original tool has delineated responses for nine of the questions: 'yes', 'no' or

Box 2. The questions in our modified CASP qualitative checklist tool

1. Was there a clear statement of the aims of the research?
 - What was the goal of the research
 - Why it was thought important
 - Its relevance
2. Is a qualitative methodology appropriate?
 - If the researcher seeks to interpret or illuminate the actions and/or subjective experiences of research participants
 - Is qualitative research the right methodology for addressing the research goal
3. Was the research design appropriate to address the aims of the research?
 - If the researcher has justified the research design (e.g. have they discussed how they decided which method to use)
4. Are the study's theoretical underpinnings (e.g. ontological and epistemological assumptions; guiding theoretical framework(s)) clear, consistent and conceptually coherent?
 - To what extent is the paradigm that guides the research project congruent with the methods and methodology, and the way these have been described?
 - To what extent is there evidence of problematic assumptions about the chosen method of data analysis? e.g. assuming techniques or concepts from other method (e.g. use of data saturation, originating in grounded theory) apply to chosen method (e.g. Braun and Clarke's reflexive thematic analysis^{39,40}) without discussion or justification.
 - To what extent is there evidence of conceptual clashes or confusion in the paper? e.g. claiming a constructionist approach but then treating participants' accounts as a transparent reporting of their experience and behaviour.
5. Was the recruitment strategy appropriate to the aims of the research?
 - If the researcher has explained how the participants were selected
 - If they explained why the participants they selected were the most appropriate to provide access to the type of knowledge sought by the study
 - If there are any discussions around recruitment (e.g. why some people chose not to take part)
6. Was the data collected in a way that addressed the research issue?
 - If the setting for the data collection was justified
 - If it is clear how data were collected (e.g. focus group, semi-structured interview etc.)
 - If the researcher has justified the methods chosen
 - If the researcher has made the methods explicit (e.g. for interview method, is there an indication of how interviews are conducted, or did they use a topic guide)
 - If methods were modified during the study. If so, has the researcher explained how and why
 - If the form of data is clear (e.g. tape recordings, video material, notes etc.)
 - If the researcher has discussed saturation of data
7. Has the relationship between researcher and participants been adequately considered?
 - If the researcher critically examined their own role, potential bias and influence during (a) formulation of the research questions (b) data collection, including sample recruitment and choice of location
 - How the researcher responded to events during the study and whether they considered the implications of any changes in the research design
8. Have ethical issues been taken into consideration?
 - If there are sufficient details of how the research was explained to participants for the reader to assess whether ethical standards were maintained
 - If the researcher has discussed issues raised by the study (e.g. issues around informed consent or confidentiality or how they have handled the effects of the study on the participants during and after the study)
 - If approval has been sought from the ethics committee
9. Was the data analysis sufficiently rigorous?
 - If there is an in-depth description of the analysis process
 - If thematic analysis is used. If so, is it clear how the categories/themes were derived from the data
 - Whether the researcher explains how the data presented were selected from the original sample to demonstrate the analysis process
 - If sufficient data are presented to support the findings
 - To what extent contradictory data are taken into account
 - Whether the researcher critically examined their own role, potential bias and influence during data analysis and selection of data for presentation
10. Is there a clear statement of findings?
 - If the findings are explicit
 - If there is adequate discussion of the evidence both for and against the researcher's arguments
 - If the researcher has discussed the credibility of their findings (e.g. triangulation, respondent validation, more than one analyst)
 - If the findings are discussed in relation to the original research question
11. How valuable is the research?
 - If the researcher discusses the contribution the study makes to existing knowledge or understanding (e.g. do they consider the findings in relation to current practice or policy, or relevant research-based literature)
 - If they identify new areas where research is necessary
 - If the researchers have discussed whether or how the findings can be transferred to other populations or considered other ways the research may be used

‘can’t tell’. Each question has a ‘comments’ box to record the reasons for a given response. The final question is open-ended. Each question is accompanied by suggested ‘hints’ to remind the researcher why the question is important. For example, the question ‘was the data collected in a way that addressed the research issue?’ is followed by seven hints. One reads: ‘(consider) If the setting for the data collection was justified’.³⁰

We reasoned that the given response categories and hints are potentially problematic for quality appraisal. First, we found that there was often more complexity and nuance to our answers than the fixed response options allowed for, including description of methodological aspects of a study that the primary authors had addressed well and also less well. It was therefore not a straightforward task to balance the methodological strengths and limitations to assign a ‘yes’ (i.e. higher quality) or ‘no’ (i.e. lower quality) response, or an especially meaningful task as a considerable amount of content and information can be lost in the process.

Moreover, it was not always clear whether the quality issue was due to methodology or reporting. For example, two papers reported one study, having used different data analysis methods to answer different research questions, and were published in a health psychology journal and an evidence-based medicine journal, which have varying disciplinary reporting conventions and publication requirements. The quality of the reporting in these papers differed to the extent that certain shared aspects of the study (i.e. modified tool questions 5 [recruitment strategy], 6 [data collection] and 7 [researcher-participant relationship]) were assigned different response options when appraised (Table 1). Furthermore, in instances when it was unclear whether the issue was methodological or one of reporting, it was also unclear whether the appropriate response was ‘no’ or ‘can’t tell’.

We agreed that there was need for a greater degree of nuance in the tool’s response options. We modified the tool by adding a fourth response option – ‘somewhat’ –, meaning ‘to some extent’ or ‘partly’, for use when we deemed that the primary study authors had reported a reasonable attempt at fulfilling a particular quality domain, but had clear strengths *and* limitations. Our quality appraisal involved assigning a total of 80 response categories across the eight papers, of which 19 (24%) were ‘somewhat’. We nominated to use ‘can’t tell’ when there was insufficient information reported to make a judgement (i.e. a reporting issue) and ‘somewhat’ when relevant information including both strengths and limitations was provided (i.e. a methodological issue).

Second, we reasoned that the hints could be (mis)interpreted as a ‘tick box’ checklist of methodological quality indicators. When interpreting the questions, we

used the hints as a means to better understand the questions and what is needed from a study to sufficiently address the questions. There is the risk that this may ‘tip over’ into inappropriately using the hints as an inadvertent quality checklist. For example, a study ‘ticks the boxes’ on four of five hints and is thus interpreted to have ‘yes’ fulfilled a particular quality domain. Prescriptive application would be problematic for two reasons. First, the number of hints per question differs (ranging from one to seven hints), meaning that questions with comparatively more hints may be ‘over-assessed’. Second, the hints are an incomplete list of quality indicators for the broad spectrum of qualitative research methods and would not apply universally. There are hints intended for specific qualitative methods of data collection (e.g. interviews), data analysis (e.g. thematic analysis) and concepts believed by some to be indicative of quality in particular methods (e.g. data saturation), but which would not be appropriate considerations for studies employing different methods and should not be considered universally applicable.

Using quality appraisal results to inform the synthesis

As per our systematic review protocol (PROSPERO registration number CRD42017083404), we used our quality appraisal results to inform the synthesis. It is generally agreed that quality appraisal results should be used meaningfully to inform a qualitative evidence synthesis.^{10,11,33,41} These uses can include the exclusion of ‘low’ quality studies from the synthesis, sensitivity analysis to examine synthesis findings with and without the inclusion of lower quality study findings, and weighting the synthesis based on quality.

In order to organise the contribution of studies based on their quality, it is necessary to distinguish studies of relative higher and lower quality. Researchers typically quantify appraisal results to produce an overall quality score for each study.¹⁸ However this is at odds with Cochrane recommended practice to avoid scoring completely and to use quality assessments to sort studies into one of three quality categories (i.e. high, medium or low).⁸ Others have agreed ‘essential’ criteria that studies must meet,^{17,42} assessed the descriptive ‘thinness’ and ‘thickness’ of evidence⁴³ and suggested establishing ‘fatal flaws’ in the design and conduct of qualitative research.³⁷ However, there is little academic consensus or empirical evidence on core criteria.

We used the CASP tool to formulate and facilitate our quality appraisal. Used as it is the tool does not necessarily produce results that are easily classified as overall ‘high’, ‘medium’ or ‘low’ quality. That is, the instrument is necessary but not sufficient for

Table 1. Comparison of quality appraisal results for two papers reporting findings from one study.

CASP question										
	1. Was there a clear statement of the aims of the research?	2. Is a qualitative methodology appropriate?	3. Was the research design appropriate to address the aims of the research?	4. Are the study's theoretical underpinnings clear, consistent and conceptually coherent?	5. Was the recruitment strategy appropriate to the aims of the research?	6. Was the data collected in a way that addressed the research issue?	7. Has the relationship between researcher and participants been adequately considered?	8. Have ethical issues been taken into consideration?	9. Was the data analysis sufficiently rigorous?	10. Is there a clear statement of findings?
Paper 1 (published in a health psychology journal)	Yes	Yes	Yes	No	Yes	Yes	Yes	Somewhat	Somewhat	Yes
Paper 2 (published in an evidence-based medicine journal)	Yes	Yes	Yes	No	Can't tell	Can't tell	Somewhat	Somewhat	Can't tell	Yes

Note: Question 11 (formerly question 10) of the CASP tool is open-ended and is therefore not included in Table 1.

distinguishing the highest and lowest quality studies. However, organising the synthesis based on quality is contingent on this step. Relatedly, sensitivity analysis on synthesis findings is performed based on specified quality appraisal criteria.^{17,42} A 'deciding criteria' is used to establish relative study quality. Review teams decide the 'tipping point' criteria based on what they consider to be important for their review aims and context.

Our systematic review and qualitative evidence synthesis was conceptualised partly in response to an inconsistent evidence base on the psychological impact of having a false positive breast screening test result. It was the first of its kind. It was therefore important for our review that the primary studies had performed robust and comprehensive data analysis, to maximise the contribution our review could make to current understanding. Further, an objective of our review was to identify areas for future research. To do this, we needed to be confident that the recommendations we made were grounded in adequate data. As a result, we considered studies in relation to the tool's criterion on the rigour of data analysis (i.e. question 9 in the modified tool) while bearing in mind how trustworthy and credible the study findings appeared as a result.⁴⁴ This, with consideration of the wider appraisal results, enabled us to identify those studies for which overall quality was relatively higher and lower.

We used our quality appraisal results to organise our thematic synthesis based on quality. In practice, this involved four additional steps to those described by Thomas and Harden for thematic synthesis.³⁵ First, we thoroughly coded the findings from studies deemed to be of higher quality to generate the main code list or coding framework. Second, the code list was used to code the findings from studies considered to be of medium quality. When it appeared relevant, appropriate or meaningful to, given our research aim and what we had learnt from the findings of higher quality studies, we created new codes based on the findings of medium quality studies. These codes were incorporated into the code list. Third, the revised code list was used to code the findings from studies determined to be of lower quality, but no new codes were generated at this point. In effect, the findings from the lower quality studies are used to support and validate the findings from higher quality studies. At this point, descriptive themes were developed from the final code list and, from there, analytical themes. Lastly, we presented verbatim extracts from all studies when reporting our synthesis. To the best of our knowledge, this particular approach has not been previously proposed.

Discussion

Overview of findings

The process of optimising the value of the CASP tool for quality appraisal in qualitative evidence synthesis has enabled us to examine several issues. We have suggested that it is necessary to appraise a qualitative primary study's approach to inquiry and we have developed an additional CASP tool question to facilitate this. We have explored how the tool's given response options and hints may influence how the tool is interpreted and affect usability. We have proposed pragmatic tool modifications to improve its suitability and usability. We have described a novel approach to organising the synthesis to prioritise findings from higher quality studies without losing relevant findings from lower quality studies.

Relevance to existing literature

Clearly reporting a study's chosen approach to inquiry for qualitative health research has been recognised as good practice. Over a decade ago, Dixon-Woods et al. acknowledged "the execution of a qualitative research study type is crucially related to the theoretical perspective in which the researchers have chosen to locate the study." (33 p. 225). More recently, the American Psychological Association (APA) has formally recognised transparent reporting of this to be best practice through their journal article reporting standards (JARS) for qualitative research in psychology.⁴⁵ The APA JARS have since been adopted as best practice guidance by other health psychology journals (e.g. British Journal of Health Psychology).⁴⁶ The British Psychological Society have published guidance for qualitative psychologists writing for the next REF exercise in the UK, which states that "general good practice guidelines often overlooked include some coherent and congruent explanation of your particular epistemological and ontological stance [...] and evidence of clear engagement with and knowledge of your approach" (15 p. 7). The development of an appropriate, additional CASP tool question to support such appraisal is thus timely and warranted. Furthermore, Sandelowski has observed that "the quality checklists that have been developed and increasingly used allow researchers to claim they have addressed quality without actually having to resolve the problems these tools raised" (47 p. 88). The addition of a question of this nature may help to improve the CASP tool's sensitivity to theoretical validity.³⁴

Our novel question is comparable to questions in the JBI critical appraisal tool.⁴⁸ Five of 10 questions in the JBI tool prompt the reviewer to consider the congruity

between the research methodology and a particular aspect of the research design. The tool has been commended for its focus on congruity.³⁴ As discussed, questions of this nature may require a certain amount of knowledge of and experience with qualitative research methods. Our CASP tool modifications are intended to preserve the strengths of the original CASP tool³³ and draw on the strengths of other existing tools, while remaining suitable for novice researchers.

Used as it is, the CASP tool may not produce results that can be used for the purposes of conducting sensitivity analysis, weighting or organising based on quality. In qualitative evidence synthesis, unlike in systematic reviews of quantitative research, decisions on essential quality criteria for these purposes are necessarily subjective. There is no empirical evidence or academic consensus on which to base such decisions. As such, qualitative evidence synthesis teams apply different quality criteria. We prioritised the tool's criteria on the rigour of data analysis to gauge the trustworthiness in the study findings in order to categorise the studies based on quality and organise the synthesis accordingly. Recent studies that used appraisal results to inform sensitivity analyses of synthesis findings applied quality of reporting criteria only¹⁷ and criteria on reporting, reflexivity and the plausibility and coherence of results.⁴² These are examples of possible criteria to apply and do not necessarily represent the 'right' or only choices of criteria; review teams should reflect on what is most valued for the aims of their review and be explicit about what they are prioritising.

Our method for organising the synthesis is comparable to published sensitivity analyses on synthesis findings, which found that lower quality studies contributed comparatively less to a synthesis compared to higher quality studies.^{17,35} We have engineered a synthesis method whereby higher quality studies contribute relatively more than lower quality studies. In this respect, our method is akin to a 'prospective' sensitivity analysis as it works on the assumption that lower quality study findings contribute relatively less to a synthesis. However, we did not assess the impact of including and excluding lower quality studies on the synthesis findings, which we understand to be a hallmark of sensitivity analysis.

The authors of reviews incorporating sensitivity analyses concluded that excluding lower quality studies did not affect their synthesis findings in any meaningful way^{17,35,42,43} and believe there is an increasingly strong case for excluding lower quality studies from a synthesis.¹⁷ However, Garside has given an alternative explanation. Despite being 'low' quality studies, the findings were evidently similar to those of higher quality studies. Including these studies "may have strengthened the

evidence for particular findings in the synthesis, especially if they had emanated from different research disciplines or had used different methods from those included in the review” (¹⁰ p. 75). Carroll et al. found potential for this to be the case. In their qualitative evidence synthesis of experiences of online learning among health professionals, studies of nurses performed less well against quality criteria compared to studies of doctors.¹⁷ If the nursing studies had been excluded, the nurse perspective would have been lost and the synthesis findings would have been based almost entirely on doctors’ experiences, thus reducing the transferability of findings. Used as it is, the CASP tool may promote an approach to appraisal that prioritises quantification of quality over content, leading to some questionable interpretations of quality. It is may not be as simple as the more ‘yes’s’ a study is assigned during appraisal, the higher the study quality. If these judgements are then used to analyse and subsequently exclude lower quality studies from a synthesis, this may be problematic for the findings and conclusions. Organising a synthesis based on quality offers an approach whereby all studies contribute to the synthesis, but the impact of lower quality studies on the synthesis findings can be moderated.

Limitations

Our findings and conclusions are based on a single case study in which the CASP tool was used to appraise studies that used semi-structured interview methods only. It is likely that we would have encountered different issues had our search identified other study types with different data types, and had we employed an alternative method of synthesis (e.g. meta-narrative synthesis or critical interpretive synthesis). We have proposed an approach to organising a synthesis based on quality that was appropriate for our mixed-quality dataset and chosen method of synthesis, but which may be less feasible for review teams depending on the context of their review.

Implications

We have described a novel approach to using appraisal results to inform the synthesis. This technique can be introduced at the first stage of Thomas and Harden’s approach to thematic synthesis without disrupting the subsequent synthesis procedure. The selection and prioritisation of one or more of the tool’s criteria, or of alternative quality principles,^{23,24} are possible means through which to understand the relative quality between studies and establish which quality category studies belong to, in order to organise the synthesis accordingly. The chosen criteria can be specified,

justified and potentially assessed in terms of the impact on the synthesis. This process can be transparent and systematic. Researchers need to be sensitive to different subpopulations, professional groups or disciplines that may receive less weighting on quality grounds as a result of adhering to different disciplinary and reporting conventions.^{10,17} Future research could further examine the process and results of organising a qualitative evidence synthesis based on quality as well as the extent to which such structuring could affect the transferability of synthesis findings.

A qualitative evidence synthesis team need experience and expertise across the qualitative methodological spectrum. A robust quality appraisal relies on broad and technical knowledge across qualitative research methods, but also on subjective judgement and tacit knowledge, which is built and shaped through experience of qualitative research. Researchers without much qualitative experience may have difficulty realising quality issues without expert help or believe that by applying the tool as it is they are performing a robust quality appraisal, when in reality it may be unsound. As we have highlighted, treating the hints as universal to all qualitative methods limits the utility of the checklist. Moreover, certain hints may restrict novice researchers’ perceptions of what constitutes good quality qualitative research because they do not apply to all methods, thus excluding other methods from the category of ‘good quality qualitative research’. The CASP tool is advocated as a suitable tool for novice researchers (although not exclusively), yet they may need further support and possibly training to apply the tool. Novice researchers will need particular support to apply our novel CASP tool question.

While it is believed by some to be good practice to report the chosen approach to inquiry, it is important to recognise that this is not necessarily common practice for all qualitative health and social care researchers. A review team can only appraise what has been reported and absence of detail is not in and of itself an indicator of poorer quality. However, in cases where an approach to inquiry has not been explicitly reported, there may be other indications of the underlying paradigm and these could be appraised for clarity, consistency and conceptual coherence, e.g. discussion of certain methodological decisions, how the data is treated and analysed, or how participants’ accounts and the findings are reported and interpreted.

At times it was not clear whether a quality issue was due to methodology or reporting. As discussed, the process of appraising quality is limited by the contents of the published report, which in turn are often limited by the publication outlet’s publishing requirements. Trying to assess quality on the basis of an inadequately reported study is problematic. As demonstrated,

inadequate reporting can affect the decisions made using the CASP tool. There have been recent calls for improved reporting of qualitative research.⁴⁹ In response, qualitative study reporting standards have been developed, such as APA JARS (for qualitative psychology research),⁴⁵ Consolidated criteria for Reporting Qualitative research (COREQ; for qualitative interview and focus group studies)⁴⁹ and Enhancing transparency in reporting the synthesis of qualitative research (ENTREQ; for systematic reviews of qualitative research).⁵⁰ Such standards represent a step in the right direction and have seemingly been well-received: the APA JARS by Levitt and colleagues was the most frequently downloaded APA article of 2018.⁵¹ These standards explain the common misconceptions of qualitative research and principles to guide appropriate evaluation, which may help journal editors and reviewers to understand and recognise the value of qualitative research submissions. Other steps may also be needed, such as extended page and word limits for qualitative and mixed-methods research reporting different study elements to those studies for which the limits were originally developed.

However, better reporting should not be seen as a proxy for quality and may not be sufficient to improve quality. Increasing open science practices offers a possible route to solution. Making the research process (e.g. study protocol, data analysis, peer review) accessible through open access publishing, open data and open data analysis workflows may increase transparency, minimise reporting bias and facilitate quality appraisal exercises.⁵² Careful consideration must be given to the appropriateness of certain open science practices for primary qualitative research, such as open data sharing and the ethical, confidentiality and sensitivity issues of this.⁵³ While there exists debates as to the extent to which such open science practices may be possible in primary qualitative research,⁵³ researchers could consider at the earliest stage of research design how and where open practices might be appropriate in their research.

Conclusions

Following our reflections on the suitability and usability of the CASP tool, we have proposed pragmatic modifications to the tool including the addition of a novel question and a novel response option. Assessing quality using the CASP tool enabled us to organise our synthesis based on quality, such that higher quality study findings contributed relatively more to the synthesis compared to lower quality study findings. Using structured approaches such as the CASP tool for the quality appraisal of qualitative research can be of value, but it is important that these

tools are applied appropriately. Novice researchers may need particular support when appraising the quality of qualitative research and synthesis teams may stand to benefit greatly from including qualitative expertise.

Acknowledgements

HAL would like to thank her PhD continuation viva examiners Professor Caroline Sanders and Dr Fiona Ulph for encouraging HAL to pursue this unexpected line of inquiry during her first PhD study (a systematic review and qualitative evidence synthesis).

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: HAL is funded by a National Institute for Health Research (NIHR) Biomedical Research Centre (BRC) PhD studentship in Manchester (IS-BRC-1215-20007). DPF is supported by the NIHR BRC in Manchester. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

ORCID iD

Hannah A Long  <https://orcid.org/0000-0001-7306-8987>

References

1. Higgins J and Green S. Chapter 22: overview of reviews. In: *Cochrane handbook for systematic reviews of interventions*. *Cochrane Database Syst Rev* 2008; 5: 187–235. Available from: <http://onlinelibrary.wiley.com/doi/10.1002/9780470712184.fmatter/summary>
2. Noyes J, Booth A, Cargo M, et al. Cochrane qualitative and implementation methods group guidance series – paper 1: introduction. *J Clin Epidemiol* 2018; 97: 35–38.
3. Hannes K, Booth A, Harris J, et al. Celebrating methodological challenges and changes: reflecting on the emergence and importance of the role of qualitative evidence in cochrane reviews. *Syst Rev* 2013; 2: 84.
4. Noblit GW and Hare RD. *Meta-ethnography: synthesizing qualitative studies*. Volume 11. Newbury Park, CA: SAGE Publications Ltd, 1988.
5. Malterud K. The impact of evidence-based medicine on qualitative metasynthesis: benefits to be harvested and warnings to be given. *Qual Health Res* 2019; 29: 7–17.
6. Thorne S. Metasynthetic madness: what kind of monster have we created? *Qual Health Res* 2017; 27: 3–12.

7. Booth A. Harnessing energies, resolving tensions: acknowledging a dual heritage for qualitative evidence synthesis. *Qual Health Res* 2019; 29: 18–31.
8. Noyes J, Booth A, Flemming K, et al. Cochrane qualitative and implementation methods group guidance series – paper 3: methods for assessing methodological limitations, data extraction and synthesis, and confidence in synthesized qualitative findings. *J Clin Epidemiol* 2018; 97: 49–58.
9. Sandelowski M and Barroso J. Reading qualitative studies. *Int J Qual Methods* 2002; 1: 74–108.
10. Garside R. Should we appraise the quality of qualitative research reports for systematic reviews, and if so, how? *Innovation* 2014; 27: 67–79.
11. Carroll C and Booth A. Quality assessment of qualitative evidence for systematic review and synthesis: is it meaningful, and if so, how should it be performed? *Res Synth Methods* 2015; 6: 149–154.
12. Yardley L. Dilemmas in qualitative health research. *Psychol Heal* 2000; 15: 215–228.
13. Crowe SF and Watt S. Excellence in research in Australia 2010, 2012, and 2015: the rising of the curate's soufflé? *Aust Psychol* 2017; 52: 503–513.
14. Franceschini F and Maisano D. Critical remarks on the Italian research assessment exercise VQR 2011–2014. *J Informetr* 2017; 11: 337–357.
15. Qualitative Methods in Psychology REF Working Group. Writing for the research excellence framework 2021: guidance for qualitative psychologists. British Psychological Society, [ww1.bps.org.uk/system/files/user-files/Qualitative. Methods in Psychology Section/public/Writing for the REF.pdf](http://www.bps.org.uk/system/files/user-files/Qualitative_Methods_in_Psychology_Section/public/Writing_for_the_REF.pdf) (2018, accessed 20 July 2020).
16. Hannes K and Macaitis K. A move to more systematic and transparent approaches in qualitative evidence synthesis: update on a review of published papers. *Qual Res* 2012; 12: 402–442.
17. Carroll C, Booth A and Lloyd-Jones M. Should we exclude inadequately reported studies from qualitative systematic reviews? An evaluation of sensitivity analyses in two case study reviews. *Qual Health Res* 2012; 22: 1425–1434.
18. Boeije HR, Van Wesel F and Alisic E. Making a difference: towards a method for weighing the evidence in a qualitative synthesis. *J Eval Clin Pract* 2011; 17: 657–663.
19. Harris JL, Booth A, Cargo M, et al. Cochrane qualitative and implementation methods group guidance series – paper 2: methods for question formulation, searching, and protocol development for qualitative evidence synthesis. *J Clin Epidemiol* 2018; 97: 39–48.
20. Cargo M, Harris J, Pantoja T, et al. Cochrane qualitative and implementation methods group guidance series – paper 4: methods for assessing evidence on intervention implementation. *J Clin Epidemiol* 2018; 97: 59–69.
21. Harden A, Thomas J, Cargo M, et al. Cochrane qualitative and implementation methods group guidance series – paper 5: methods for integrating qualitative and implementation evidence within intervention effectiveness reviews. *J Clin Epidemiol* 2018; 97: 70–78.
22. Flemming K, Booth A, Hannes K, et al. Cochrane qualitative and implementation methods group guidance series – paper 6: reporting guidelines for qualitative, implementation, and process evaluation evidence syntheses. *J Clin Epidemiol* 2018; 97: 79–85.
23. Dalton J, Booth A, Noyes J, et al. Potential value of systematic reviews of qualitative evidence in informing user-centered health and social care: findings from a descriptive overview. *J Clin Epidemiol* 2017; 88: 37–46.
24. Cho J and Trent A. Validity in qualitative research revisited. *Qual Res* 2006; 6: 319–340.
25. Elliott R, Fischer CT and Rennie DL. Evolving guidelines for publication of qualitative research studies in psychology and related fields. *Br J Clin Psychol* 1999; 38: 215–229.
26. Levitt HM, Motulsky SL, Wertz FJ, et al. Recommendations for designing and reviewing qualitative research in psychology: promoting methodological integrity. *Qual Psychol* 2017; 4: 2–22.
27. Smith B and McGannon KR. Developing rigor in qualitative research: problems and opportunities within sport and exercise psychology. *Int Rev Sport Exerc Psychol* 2018; 11: 101–121.
28. Yardley L. Demonstrating validity in qualitative psychology. In: Smit JA (ed.) *Qualitative psychology: a practical guide to research methods*. London: SAGE, 2015: 257–272, chapter 12.
29. Munthe-Kaas HM, Glenton C, Booth A, et al. Systematic mapping of existing tools to appraise methodological strengths and limitations of qualitative research: first stage in the development of the CAMELOT tool. *BMC Med Res Methodol* 2019; 19: 1–13.
30. Critical Appraisal Skills Programme. CASP qualitative checklist, <https://casp-uk.net/wp-content/uploads/2018/01/CASP-Qualitative-Checklist-2018.pdf> (accessed 14 February 2020).
31. Hannes K and Bennett S. Understanding evidence from qualitative research. In: Hoffman T, Bennett S and Del Mar C (eds) *Evidence-based practice across the health professions*. 3rd ed. Chatswood, NSW: Elsevier Australia, 2017, pp.226–247.
32. Long H, Brooks JM, Harvie M, et al. How do women experience a false-positive test result from breast screening? A systematic review and thematic synthesis of qualitative studies. *Br J Cancer* 2019; 121: 351–358.
33. Dixon-Woods M, Sutton AJ and Shaw RL. Appraising qualitative research for inclusion in systematic reviews: a quantitative and qualitative comparison of three methods. *J Heal Serv Res Policy* 2007; 12: 42–47.
34. Hannes K, Lockwood C and Pearson A. A comparative analysis of three online appraisal instruments' ability to assess validity in qualitative research. *Qual Health Res* 2010; 20: 1736–1743.
35. Thomas J and Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Med Res Methodol* 2008; 8: 45.

36. Johnson RB. Dialectical pluralism: a metaparadigm whose time has come. *J Mix Methods Res* 2017; 11: 156–173.
37. Dixon-Woods M, Shaw RL, Agarwal S, et al. The problem of appraising qualitative research. *Qual Saf Heal Care* 2004; 13: 223–225.
38. Sandelowski M. Rigor or rigor mortis: the problem of rigor in qualitative research revisited. *ANS Adv Nurs Sci* 1993; 16: 1–8.
39. Braun V and Clarke V. Thematic analysis. *APA Handb Res Methods Psychol Vol 2 Res Des Quant Qual Neuropsychol Biol* 2012; 2: 57–71.
40. Braun V and Clarke V. Reflecting on reflexive thematic analysis. *Qual Res Sport Exerc Heal* 2019; 11: 589–597.
41. Dixon-Woods M, Booth A and Sutton AJ. Synthesizing qualitative research: a review of published reports. *Qual Res* 2007; 7: 375–421.
42. Franzel B, Schwiegershausen M, Heusser P, et al. How to locate and appraise qualitative research in complementary and alternative medicine. *BMC Complement Altern Med* 2013; 13: 125.
43. Noyes J and Popay J. Directly observed therapy and tuberculosis: how can a systematic review of qualitative research contribute to improving services? A qualitative Meta-synthesis. *J Adv Nurs* 2007; 57: 227–243.
44. Lincoln Y and Guba E. *Naturalistic inquiry*. Beverly Hills, CA: SAGE Publications Ltd, 1985.
45. Levitt HM, Bamberg M, Creswell JW, et al. Reporting standards for qualitative research in psychology: the APA publications and communications board task force report. *Am Psychol* 2018; 1: 26–46.
46. Shaw RL, Bishop FL, Horwood J, et al. Enhancing the quality and transparency of qualitative research methods in health psychology. *Br J Health Psychol* 2019; 24: 739–745.
47. Sandelowski M. A matter of taste: evaluating the quality of qualitative research. *Nurs Inq* 2015; 22: 86–94.
48. The Joanna Briggs Institute. Checklist for qualitative research. Joanna Briggs Inst [Internet]. 2016. 6, www.joannabriggs.org/assets/docs/critical-appraisal-tools/JBI_Critical_Appraisal-Checklist_for_Qualitative_Research.pdf. (2016, accessed 20 July 2020).
49. Tong A, Sainsbury P and Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care* 2007; 19: 349–357.
50. Tong A, Flemming K, McInnes E, et al. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Med Res Methodol* 2012; 12: 181–188.
51. Winerman L. The top 10 journal articles. *Monitor on Psychology* 2018, pp.36–41, www.apamonitor-digital.org/apamonitor/201812/MobilePagedReplica.action?pm=2&folio=36#pg39 (accessed 20 February 2020).
52. Munafò MR, Nosek BA, Bishop DVM, et al. A manifesto for reproducible science. *Nat Hum Behav* 2017; 1: 1–9.
53. Branney P, Reid K, Frost N, et al. A context-consent meta-framework for designing open (qualitative) data studies. *Qual Res Psychol* 2019; 16: 483–502.