

# Applying GRADE-CERQual to Interpretive Review Findings: Reflections From a Cochrane meta-ethnography on Childhood Vaccination Acceptance

International Journal of Qualitative Methods

Volume 23: 1–15

© The Author(s) 2024

DOI: 10.1177/16094069241246413

[journals.sagepub.com/home/ijq](https://journals.sagepub.com/home/ijq)

Sara Cooper<sup>1,2,3</sup> , Natalie Leon<sup>4,5</sup>, Bey-Marrie Schmidt<sup>5,6</sup>, Alison Swartz<sup>2,7</sup>, Charles S. Wiysonge<sup>1,8</sup> , and Christopher J. Colvin<sup>2,4,9,10</sup> 

## Abstract

GRADE-CERQual (Confidence in the Evidence from Reviews of Qualitative research) was developed to support the use of evidence from qualitative reviews within policy- and decision-making. To date, the approach has been applied predominantly to aggregative synthesis methodologies and descriptive review findings. GRADE-CERQual guidance recommends the approach be tested on more diverse review methodologies and outputs to support its evolution. This paper contributes to this evolution by reflecting on our experiences of applying GRADE-CERQual to findings that emerged from a recent Cochrane meta-ethnography on childhood vaccination. Specifically, we describe the similarities and differences, challenges and dilemmas we experienced applying the approach to more interpretive versus more descriptive review findings. We found that we were able to apply the core criteria and principles of GRADE-CERQual in ways that were congruent with the methodologies and epistemologies of a meta-ethnography and its findings. We also found that the practical application processes were similar across review finding types. The main differences related to the level of demand placed on the evidence and the level of complexity involved with the decisions. Compared to more descriptive findings, more interpretive findings required evidence that was richer, thicker, more contextually situated and methodologically stronger for us to have the same level of confidence in them. Making the assessments for these findings also involved more complicated forms of judgement. We provide practical examples to illustrate these complexities and how we approached them, which others applying GRADE-CERQual to more interpretive review findings could draw upon. We also highlight areas requiring further discussion, in the hope that this will offer a platform for engagement and the potential future refinement of the approach. Ultimately, this could enhance the usability of GRADE-CERQual for a larger range of qualitative review findings and in turn expand the kinds of knowledges that count within decision-making.

## Keywords

GRADE-CERQual, qualitative evidence synthesis (QES), meta-ethnography, interpretive review findings

<sup>1</sup>Cochrane South Africa, South African Medical Research Council, South Africa

<sup>2</sup>School of Public Health and Family Medicine, University of Cape Town, South Africa

<sup>3</sup>Department of Global Health, Faculty of Medicine and Health Sciences, Stellenbosch University, South Africa

<sup>4</sup>Department of Epidemiology, School of Public Health, Brown University, USA

<sup>5</sup>Health Systems Research Unit, South African Medical Research Council, South Africa

<sup>6</sup>School of Public Health, Faculty of Community Health Sciences, University of the Western Cape, South Africa

<sup>7</sup>Institute of Medical and Biomedical Education, St George's University of London, UK

<sup>8</sup>Vaccine Preventable Diseases Programme, World Health Organization Regional Office for Africa, Congo

<sup>9</sup>Department of Public Health Sciences, University of Virginia, USA

<sup>10</sup>Global Health Center, Graduate Institute of International and Development Studies, Switzerland

## Corresponding Author:

Sara Cooper, Cochrane South Africa, South African Medical Research Council (SAMRC), Francie van Zijl Drive, Parow Valley, Cape Town 7501, South Africa.

Email: [sara.cooper@mrc.ac.za](mailto:sara.cooper@mrc.ac.za)



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

## Introduction

Over the last 15 years there has been growing recognition of the potential contribution of qualitative evidence within global health and social care decision-making (Carmona et al., 2021; Langlois et al., 2018). Those working in these arenas increasingly seek evidence beyond the effects of interventions, to wider questions about local norms and preferences, equity and human rights issues, acceptability and feasibility of interventions, implementation processes, and the impact of socio-political and cultural contexts (Flemming & Noyes, 2021; Lewin, Booth, et al., 2018). Qualitative research, and particularly reviews of qualitative evidence, are increasingly seen to offer important insights for answering this broader range of questions (Lewin & Glenton, 2018).

‘Qualitative evidence syntheses’ (QES) - or systematic reviews of qualitative evidence – is a term for the broad group of methods for systematically synthesising the findings from multiple primary qualitative studies (Noyes et al., 2018a). QES methods tend to follow a similar logic to a quantitative systematic review, however, their procedures are tailored to the significant methodological and epistemological differences between quantitative and qualitative research (Hannes & Macaitis, 2012). QES has recently become an important method for incorporating qualitative research into health and social care decision-making processes, including global guideline development and policy formulation. For example, over the last decade various World Health Organisation (WHO) guidelines have included findings from QES to determine what outcomes were important to stakeholders, or to inform the values and preferences, acceptability, feasibility, and/or equity criteria of the respective evidence-to-decision (EtD) frameworks (Downe et al., 2019; Glenton et al., 2019; Lewin et al., 2019). It has indeed been suggested that the growing recognition and use of qualitative research within decision-making means we may be “entering a new era” for qualitative research (Lewin & Glenton, 2018).

It is against this backdrop that the GRADE (Grading of Recommendations Assessment, Development and Evaluation)-CERQual (Confidence in the Evidence from Reviews of Qualitative research) approach was developed to support the use of findings from QES in decision-making (Lewin, Booth, et al., 2018; Lewin et al., 2015). GRADE-CERQual provides a systematic and transparent framework for assessing how much confidence decision-makers and other users can place in individual review findings from QES. ‘Confidence’ is understood as an assessment of the extent to which a review finding is a reasonable representation of the phenomenon of interest (Lewin, Booth, et al., 2018). The GRADE-CERQual approach complements and shares similar objectives to GRADE tools for other types of evidence (Guyatt, Oxman, Akl, et al., 2011;

Hsu et al., 2011; Lewin, Booth, et al., 2018). However, it is based on principles and concepts of qualitative research and was designed specifically for application in a QES. Authors of QES are increasingly incorporating GRADE-CERQual assessments in their reviews as a marker of best practice (Flemming & Noyes, 2021). The most up-to-date guidance on applying the approach is available as a special series of articles published 2018 in Implementation Science (Lewin, Booth, et al., 2018).

To date, however, GRADE-CERQual has mainly been applied to evidence syntheses that have used more aggregative analysis methods and that have produced largely descriptive findings (Bohren et al., 2023; Wainwright et al., 2023). There is much less experience with applying the approach to more interpretive findings, such as broader concepts, logic models or theory, that may emerge from more interpretive synthesis methodologies (Brookfield et al., 2019; Flemming & Noyes, 2021; Noyes et al., 2018b). The aspiration is that the approach could be applied to any type of qualitative review finding and synthesis method (Lewin, Booth, et al., 2018). There is therefore a need to test the approach with a wider range of qualitative review findings and methods to assess whether it may need to be expanded or adapted (Wainwright et al., 2023). Indeed, GRADE-CERQual is currently conceptualised as an emerging approach, and it is anticipated that guidance will evolve over time as experience is gained on its application across more diverse review findings and synthesis approaches (Glenton et al., 2018).

In this paper we seek to contribute to this evolution by reflecting on our experiences of applying GRADE-CERQual to the review findings that emerged from a recent Cochrane meta-ethnography we conducted on childhood vaccination acceptance (Cooper et al., 2021). Specifically, we describe both the similarities as well as the differences, challenges and dilemmas we experienced when applying the approach to the more interpretive review findings compared to the more descriptive review findings. We provide practical examples to illustrate the complexities we faced and how we approached them, which others applying GRADE-CERQual to more interpretive review findings could draw upon. Our experience also generated various questions, which we reflect upon in this paper and flag for greater thought and discussion. Our hope is that this can provide a platform for further engagement on these issues, and the potential future refinement of guidance on applying GRADE-CERQual.

We recognize and share some of the concerns within more critical qualitative research communities about the growing use of qualitative research within policy- and decision-making (Lambert et al., 2006; Mykhalovskiy & Weir, 2004; Sandelowski et al., 1997; Thorne et al., 2004), as further unpacked in the conclusion of this paper. Yet we believe that enhancing the usability of GRADE-CERQual for a wider range of qualitative research findings and methodologies holds significant transformative potential for expanding the kinds of knowledges and ways of knowing that count.

## *The Cochrane Meta-Ethnography on Childhood Vaccination Acceptance: Methods and Types of Review Findings*

A detailed description of the methods and findings of our review are reported elsewhere (Cooper et al., 2021). In summary, our review sought to develop a conceptual understanding of what and how different factors interact to influence parental views and practices around routine childhood vaccination. We used a meta-ethnographic approach for the synthesis, drawing heavily on the analytical steps outlined originally by Noblit and Hare (Noblit & Hare, 1988) and the eMERGe meta-ethnography reporting guidance (France, Cunningham, et al., 2019). Meta-ethnography is an interpretive (as opposed to aggregative) qualitative synthesis approach which translates and synthesises conceptual data from included studies to produce more interpretive or higher-level understandings.

Using this approach, we produced various types of review findings. In particular, and in line with Sandelowski and Barroso (Sandelowski & Barroso, 2007), we conceived qualitative findings as existing along a spectrum of data transformation. On the one end of the spectrum are more descriptive findings, which describe patterns in the data. On the other end of the continuum are more interpretive or explanatory review findings, which provide theoretical interpretations or explanations of the patterns in the data. That is, descriptive review findings essentially name or describe a phenomenon, whereas interpretive review findings make claims about how that phenomenon is produced or acts upon the world. Typical descriptive findings in our review included, for example, findings about the influence on vaccine acceptance of ‘religious beliefs’ or ‘access challenges’ or ‘distrust in expert systems’. More interpretive findings from our review comprised, for example, findings related to how social communities and vaccination views exist in a mutually reinforcing relationship, and how phenomena such as ‘social exclusion’ and ‘neoliberalism’ constitute potential pathways for reducing vaccination acceptance.

We recognise, however, that this distinction rings both true and false in important ways. Labelling one review finding as ‘descriptive’ and another as ‘interpretive’ inevitably misrepresents what is essentially a continuum of review finding types. Moreover, all types of review findings are arguably interpretations, inevitably constructed through the interpretive lens of the review authors. In our review we therefore used this distinction for the utility it served, whilst simultaneously appreciating the inherent problems with its usage.

## **Findings**

In this section we reflect on our experience of operationalising GRADE-CERQual, including developing ‘summary of findings’ and applying each of the four components of the approach. In particular, we focus on both the similarities as well as the differences, challenges and dilemmas we faced when applying GRADE-CERQual to our more interpretive versus

more descriptive findings. Throughout we flag various questions and topics which we think could benefit from further scrutiny and debate. In Tables 1 and 2 we provide an example from our review of a finding, ‘summary of finding’ and GRADE-CERQual assessments of a more descriptive and a more interpretive review finding respectively. We refer to these examples throughout to help illustrate the different issues we faced and how, at times, we approached them.

### *Developing ‘Summary of Findings’*

The first step when applying GRADE-CERQual involves developing a short statement or ‘summary of finding’ that provides a succinct, but clear, description of each review finding (Lewin, Bohren, et al., 2018). The GRADE-CERQual assessments are then applied to each individual ‘summary of finding’, which in turn form the basis of the Evidence Profile and Summary of Qualitative Findings (SoQF) tables (Lewin, Bohren, et al., 2018).

Developing ‘summary of findings’ was relatively straightforward for the more descriptive findings, as both their construction and meaning were usually fairly simple. For example, our more descriptive finding on ‘Socio-economic challenges in accessing vaccination services’ (Table 1) provides a relatively straightforward report of the barriers parents face in obtaining vaccination and the different ways these can impact on acceptance of vaccines. Translating this finding into a ‘summary of finding’ was therefore an uncomplicated task. In contrast, crafting ‘summary of findings’ for our more interpretive findings was a lot more challenging. As depicted in our finding on a ‘neoliberal logic’ (Table 2), most of our more interpretive findings were relatively complex conceptual abstractions, and therefore it was not always clear how these might be summarised into more useable statements. They also tended to have many different component concepts, often with varying definitions and accompanying theories. We did not necessarily want to incorporate all the component concepts and theories, and we needed to unpack what we were (and were not) meaning by our use of the terms. All of this obviously requires a fair degree of explanation to the reader, something which is challenging to capture in a short statement.

Therefore, unlike with our more descriptive findings, for our more interpretive findings we decided that we needed some explicit principles to guide the crafting of ‘summary of findings’. Here we agreed on two principles. Firstly, that the goal was to distil out the fundamental mechanism at work for each finding. That is, the objective was not to try and incorporate all the layers and parts of the finding, but rather to capture the core issue that connected the different threads of the finding. A second principle we used was to consider the end users of the review, which in our case was predominantly policymakers and healthcare practitioners. That is, we decided we needed to package the more complex interpretive findings in a potentially more useful and actionable way.

**Table 1.** Descriptive Review Finding Example: Finding, ‘Summary of Finding’ and GRADE-CERQual Assessments.<sup>a</sup>

**Finding:** Studies within resource-limited settings revealed how livelihood struggles can pose significant barriers to accessing vaccination services. Whilst parents in these contexts may hold a general belief in the value of vaccination for their children, socioeconomic challenges may reduce their acceptance of vaccination (Millimouno, 2006; Leach, 2007; Dugas, 2009; McKnight, 2014; Giles-Vernick, 2016; Smith, 2017). These challenges may include practical questions of geography and transport, childcare constraints, and family economics and household work pressures. For example, parents in studies from multiple African countries (Gambia, Guinea, Ethiopia, Central African Republic, and Burkina Faso) spoke about the long distances they have to travel (up to 40 km in some localities), often by foot, to reach the nearest site where vaccination is provided (Millimouno, 2006; Leach, 2007; Dugas, 2009; Giles-Vernick, 2016). Similar findings emerged in a study amongst site-dwelling Gypsy, Roma, and Traveller parents in Britain (Smith, 2017). These parents frequently reside in geographically- and socially- isolated locations, often poorly served by public transport and other services, which hindered their ability to access immunisation services. Many parents also highlighted the difficulties they experience finding someone to care for their other children or family members who cannot be left at home alone while they make the trip to the clinic (Millimouno, 2006; Leach, 2007; McKnight, 2014).

Parents emphasised how taking their child for vaccination requires arranging for someone to look after these other family members or enduring the additional burden of bringing them along. At the same time, many parents in these settings in Africa and Britain spoke at length about the constant, daily juggling of economic activities and domestic tasks that they have to undertake so as to keep their households afloat (Millimouno, 2006; Leach, 2007; McKnight, 2014; Giles-Vernick, 2016; Smith, 2017). As described by the authors of all these studies, vaccination constitutes one of the many basic priorities for these parents, and therefore needs to be fitted into an often already demanding daily schedule. They suggest that the day-to-day demands of living ultimately make taking time out for vaccinating a challenge for many of these parents, frequently resulting in a trade-off with another activity that is of central importance for their survival. This balancing act was clearly captured by one parent from Bangui, Central African Republic, who explained, “My children haven’t gotten all of their vaccinations... I have to work every day at the market... if I don’t go to the market, my family won’t have any money. I know that I had to bring my children to receive their vaccinations. But I couldn’t do it” (Giles-Vernick, 2016; Central African Republic, participant quote). Numerous parents provided analogous accounts to this one, similarly attributing their avoidance or delay in vaccinating their children to the time, effort, and opportunity costs that vaccination involved for them.

**Summary of finding:** “Socio-economic challenges in accessing vaccination services”: Parents living in resource-limited settings frequently face numerous socioeconomic challenges to accessing vaccination services, including practical questions of geography and transport, childcare constraints, and family economics and household work pressures. As a result of these challenges, some of these parents were less accepting of vaccination due to the time, effort, and opportunity costs that accessing vaccination involved.

#### Assessment for each GRADE-CERQual component

|                            |  |
|----------------------------|--|
| Methodological limitations | Minor concerns due to poor reporting of data analysis methods, and limited evidence of reflexivity and sensitivity to ethical concerns in most studies |
| Coherence                  | No or minor concerns (the finding reflects the variation and complexity of the data, and is well supported by details in the underlying studies)       |
| Relevance                  | Minor concerns because 5 of the 6 studies were conducted in Africa, yet 1 study was conducted in Britain   |
| Adequacy                   | Minor concerns (6 studies contributing relatively thin data)   |

#### Overall GRADE-CERQual assessment and explanation

High confidence

<sup>a</sup>Some of the details have been slightly adapted from the original qualitative evidence synthesis to illustrate certain issues regarding making GRADE-CERQual assessments.

Guided by these principles, the development of the ‘summary of findings’ for the more interpretive findings ended up most often being an additional analytical step, rather than just a matter of summarising and expressing. That is, it usually required an additional interpretive process to ‘translate’ them into something more succinct and practical. Relatedly, the ‘summary of findings’ we produced was less of a reflection of the finding per se, and more an aspect of it with a particular angle. Consequently, our ‘summary of findings’ could have focused on a different aspect.

On reflection, however, this process and the outputs produced were, in fact, not intrinsically different for our more descriptive and more interpretive review findings. The formulation of all our ‘summary of findings’, at least to some degree, involved critical reflection about the content of the full review finding. This iterative process therefore

inevitably formed part of the analysis, and at times led to refinements of findings and ‘summary of findings’ for all types of findings. Similarly, the ‘summary of findings’ across the spectrum of finding types involved choices around what to highlight and how to highlight them, although these choices were potentially less overt when the findings were more descriptive and therefore more straightforward. In other words, it was always possible to construct different ‘summary of findings’ based on the same finding and associated data. And in all cases, how we framed the ‘summary of findings’ (even small wording tweaks) gave rise to different confidence threats and in turn different GRADE-CERQual assessments. This is, indeed, routine procedure when applying GRADE-CERQual: concerns or limitations regarding the underlying evidence may be presented in the ‘summary of finding’ itself and how it is framed, or in the assessment.



**Table 2.** Interpretive Review Finding Example: Finding, 'Summary of Finding' and GRADE-CERQual Assessments.<sup>a</sup>

**Finding:** One potential pathway to reduced acceptance of childhood vaccination can be captured by the concept of 'neoliberal logic'. This concept encapsulates the interconnected and frequently coexisting ideas held by many parents, which come together to create a particular logic of care regarding vaccination. This logic of care was common amongst parents across socioeconomic divides in HICs. It also emerged amongst parents from middle- and higher-socioeconomic groups in Brazil (Barbieri, 2017). This logic can be characterised as follows. With a view of their child's health as individualised and inevitably fragile, parents attempt to separate out the risks of disease and the potential benefits of vaccines as they pertain to their specific child. With a view of themselves as experts of their children and a scepticism of the expert systems implicated with vaccination, parents perceive themselves as best placed to conduct this risk-benefit calculus for their child. With a strong conviction that they are singularly responsible for their child's health, they see this risk assessment as a moral obligation and the vaccination decision taken as ultimately a matter of personal choice.

This particular logic of care is not an 'individual' belief held by a select few. Rather, it is deeply rooted in, and reflects, trends in thinking that have become ubiquitous within contemporary, industrialised societies. Within these contexts, the ideas of individual responsibility and personal choice have become dominant values in recent decades. This is related to the processes of individualisation and a rollback of the state to a political agenda stressing citizens' own rights and responsibilities. Drawing on the works of sociologists Ulrich Beck and Anthony Giddens (Giddens, 1990; Giddens, 1991; Beck, 1992), many study authors highlight how we are currently living within a 'risk society' era (Rogers, 1995; Petts, 2004; Poltorak, 2005; Casiday, 2007; Brunson, 2013; Sobo, 2016). In this era, risks have intensified in magnitude and become increasingly global: International disease outbreaks, nuclear war, environmental pollution, global warming, accidents, terrorism, and so forth. The nature and scope of contemporary risks make them more difficult than in the past to calculate and control, creating a climate of increased vulnerability. At the same time, central institutions and systems (government, industry, science) are increasingly viewed as partly responsible for, and unable to control, the types of risks thrown up by the processes of modernisation. Distrust and suspicion of these institutions have therefore become pervasive features of contemporary social life.

Within this context, there has been an intensification of concern with assessing, avoiding, and managing risk, processes which have become increasingly individualised and understood as matters of personal responsibility and choice. Drawing on the thinking of French philosopher Michel Foucault (Foucault, 1977; Foucault, 1989; Foucault, 1991), various study authors highlighted how people are increasingly required to become 'responsibilised' citizens who self-govern their lives, remain actively aware of dangers and opportunities, and who recognise that they themselves are accountable for their own well-being (Petts, 2004; Ward, 2017). Ultimately, making decisions in terms of individual choice and responsibility has become a moral imperative and has infiltrated many domains of life.

Health and healthcare are no exception in this regard, where individual responsibility for one's own health (or that of one's children) has become a major societal value. Using the theory of "healthism", originally coined by Robert Crawford (Crawford, 1980; Crawford, 2004), various study authors demonstrate how people are encouraged to empower themselves to become "active" (Sobo, 2016), and "enterprising and entrepreneurial" (Peretti-Watel, 2019), healthcare consumers who undertake "self-initiated health promoting behavior" (Rogers, 1995). Within this neoliberal logic, health has also become a moral imperative and an essential aspect of self-worth, whereby a 'failure' to look after one's health or to make 'good' healthcare choices is seen as a moral transgression (Reich, 2016; Ward, 2017). The 'holistic' or 'natural' health movement is a product, at least in part, of this heightened preoccupation with personal health and the individual lifestyle modifications thought to promote or maintain it (Crawford, 1980; Crawford, 2004). This ideology of 'healthism' dovetails with, and is reinforced, by a contemporary public health paradigm that strongly promotes the values of 'patient choice', 'patient empowerment', 'self-management', and 'shared decision-making' in health and healthcare practices. As suggested by several study authors, health promotion policy and practice consistently urge people to be independent and self-efficacious, to educate themselves on health-related issues and to fully participate in making decisions to optimise their own (and their families') health and well-being (Rogers, 1995; Poltorak, 2005; Brunson, 2013; Reich, 2016; Sobo, 2016; Ward, 2017).

Views regarding childhood vaccination for many parents in HICs therefore need to be understood within this context and the forms of neoliberal subjectivity it has produced. These kinds of vaccination views, informed by neoliberal discourses, were common amongst parents across the spectrum of vaccination attitudes in HICs: those who accepted vaccination, those who were hesitant towards vaccination, and those who refused vaccination. However, for certain parents, these views were experienced as in conflict with vaccination promotion messages. Some of these parents were particularly uneasy with the population-level risk-benefit statistics that vaccination promotion discourses provide, which they felt disregarded their child's individual vaccination needs and vulnerabilities. Other parents resented the emphasis such discourses placed on community obligation and the health of the 'herd', when being a good parent is associated with taking personal responsibility for what's best for one's own child. Other parents resented the claim of 'doctor knows best' underpinning vaccination promotion messages, when they considered themselves to have the best knowledge of their child's strengths and vulnerabilities. Many of these parents had an inherent distrust for health authorities, and therefore objected to vaccination promotion messages asking them to be 'a blind follower of the state' and to 'uncritically trust' what health authorities promote. Some parents also pointed to and resented the inconsistencies between these messages and other health promotion discourses which advocate for independent, reflexive decision-making and personal choice.

Therefore, the source of conflict certain parents perceived between vaccination promotion messages and their views about health and healthcare informed by neoliberal discourses varied. However, what appeared common for these parents, and what potentially separated them from other parents who accepted vaccination, was that this experienced tension was absolute and unacceptable. This, in turn, may have led them to resist vaccination promotion messages and be less accepting of vaccination.

(continued)

Table 2. (continued)

|  |  |
|--|--|
| <b>Summary of finding:</b> “Neoliberal logic”: Many parents, predominantly from HICs, held a worldview informed by neoliberal discourses. These discourses understand health as individualised, and health-related risks and decisions as matters of individual choice and responsibility. Furthermore, these discourses suggest that being a ‘good’ and responsible person in the world means consistently assessing one’s own, individual health-related risks, seeking and questioning evidence about such risks, proactively avoiding and managing such risks, and understanding that one is singularly accountable for the outcomes that ensue. Some parents experienced this worldview as in conflict with vaccination promotion messages, which tend to be informed by other types of discourses. The discourses underpinning vaccination promotion messages frequently emphasise generalised risk and safety statistics, claim that ‘doctor knows best’, ask people to trust authorities and follow what they promote, and advocate for social responsibility and the health of the community. Parents’ perceived tension between the discourses informing their own worldview and those of vaccination promotion messages led some to resist these messages and be less accepting of vaccination. |  |
| <b>Assessment for each GRADE-CERQual component</b>   |  |
| Methodological limitations   | Moderate concerns due to poor reporting of data collection and analysis, and limited evidence of sensitivity to ethical concerns or reflexivity in many of the studies   |
| Coherence  | Moderate concerns: For some aspects of the finding, there was less support from the underlying data (e.g. proactively avoiding and managing risks), or it was hard to tell the level of support from the underlying data because the issue (e.g. parents’ experienced conflict) was mentioned in passing and not explored in detail by many of the studies (ambiguous data). However, the theory of neoliberalism is well established and its various components and tensions well-mapped out in existing literature. This therefore slightly increased our confidence. We also explored alternative explanations and found that the data supports various alternative theoretical interpretations (e.g. psychological theories on risk beliefs and appraisal and utility calculation). However, we did not identify any theory that contradicted or refuted our interpretation: All plausible alternatives we explored complemented our interpretation, albeit providing a different perspective or paradigm in which to interpret the findings. We therefore did not judge these plausible (compatible) alternatives to pose a serious threat to coherence. However, and in sum, this is a conceptually transformed finding that has moved somewhat away from the underlying data and the primary focus of the studies, and various inferences have been made. We therefore do have moderate concerns regarding its coherence. |
| Relevance  | No or very minor concerns (studies include a wide range of participants, settings, and vaccines within HICs)   |
| Adequacy   | No or very minor concerns (15 studies, together contributing a large amount of rich data)  |
| <b>Overall GRADE-CERQual assessment and explanation</b>  |  |
| Moderate   | Finding downgraded because of moderate concerns about coherence (ambiguous data) and moderate concerns about methodological limitations  |

<sup>a</sup>Some of the details have been slightly adapted from the original qualitative evidence synthesis to illustrate certain issues regarding making GRADE-CERQual assessments.

For example, the evidence might suggest that a preference for homeopathic interventions is more common amongst parents from high income countries (HICs). One could reflect this by writing it into the ‘summary of finding’ by indicating that “*Many parents, particularly in HICs, had a preference for homeopathic interventions*”, in which case your confidence assessment would be high. Alternatively, you could leave this out of the ‘summary of finding’ by indicating that “*Many parents had a preference for homeopathic interventions*” in which case you would lower your confidence to moderate or low.

Currently, ‘summary of findings’, along with their GRADE-CERQual assessments, are frequently the main source of evidence that is used in guideline and other decision-making processes (Bohren et al., 2023; Lewin et al., 2018) Yet there is currently little discussion, and some ambiguity, within current guidance about the relationship between the ‘summary of findings’, the full finding and the GRADE-CERQual assessment. Greater clarity on this relationship, and how it might be more explicitly reflected in the presentation of the evidence to end users, would be useful.

### Applying the four GRADE-CERQual components

Once we had developed the ‘summary of findings’ we then proceeded to make the GRADE-CERQual assessments for each individual ‘summary of finding’. GRADE-CERQual currently assesses confidence in a review finding based on four key components: the adequacy of data supporting the review finding; the relevance of the individual studies contributing to the review finding; the methodological limitations of the individual qualitative studies contributing to the review finding; and the coherence of the review finding (Bohren et al., 2023; Lewin et al., 2018).

When making the GRADE-CERQual assessments, we used the same core criteria and principles, and followed the same practical process, for the more descriptive and the more interpretive review findings. The main differences we experienced related to, firstly, the level of demand placed on the evidence and secondly, the level of complexity involved with the judgements. That is, as the review findings became more interpretive and in turn more transformed, the demands from the evidence supporting the review finding increased. In other words, for us to have ‘moderate’ or ‘high’ confidence in an

interpretive review finding required considerably more from the evidence than what was expected for ‘moderate’ or ‘high’ confidence in a more descriptive finding.

A second major difference we experienced when making the GRADE-CERQual assessments for the more descriptive versus more interpretive review findings related to the level of complexity involved with the judgements, with the latter requiring more complicated and challenging decisions. This was generally the case, but most specifically for the component of coherence. Below we illustrate these issues for each of the four GRADE-CERQual components separately.

### *Component 1: Adequacy*

We assessed the adequacy component by asking the same general question across the spectrum of review finding types: do we have sufficient data on the phenomenon of interest to feel confident about the review finding? In line with current guidance, in all cases we assessed two aspects of the adequacy component: the extent to which the information provided in the studies is detailed enough to allow the review authors to interpret the meaning and context of the phenomenon of interest (‘data richness’); and the extent to which the amount of studies and participants contributing to the review finding are adequate (‘quantity of data’) (Glenton et al., 2018). For all review findings, if we deemed there to be significant threats, we either lowered our confidence or reformulated the finding so as to strengthen our confidence in its adequacy. This is routine procedure when applying GRADE-CERQual, as demonstrated earlier-concerns or limitations regarding the underlying evidence may be presented in the ‘summary of finding’ itself and how it is framed, or alternatively in the assessment.

When assessing the adequacy component, like with all the components of GRADE-CERQual, our judgements were made *in relation* to the nature of the review finding and the claims it is making. Our more interpretative findings are all making fairly broad and complex claims about phenomena, social structures, relationships and processes. For example, our finding about a ‘neoliberal logic’ (Table 2) suggests the existence of a worldview, makes claims about the social forces producing this worldview, and proposes various mechanisms for how this worldview may lead to a reduction in vaccination acceptance. For us to be confident in the adequacy component of this complex finding, the data from contributing studies needed to be rich enough to allow for an adequate understanding of the phenomena described in the review finding, and the quantity of data needed to be sufficiently large enough to be able to support the broad claims being made. In the case of our ‘neoliberal logic’ (Table 2) finding, fifteen studies contributed to it, with ten of these studies providing detailed information about the meaning and interactions of the different factors. We therefore decided that, albeit complex, this review finding is sufficiently supported by the data and concluded that we have no or very minor concerns about data adequacy.

In contrast, our more descriptive findings tended to be narrower in scope, and the claims being made much simpler. For example, our finding related to ‘socio-economic challenges in accessing vaccination services’ (Table 1) essentially labels the barriers parents face in obtaining vaccination and reports that these impact on acceptance of vaccines. Six studies contributed to this finding, with all studies offering somewhat little or superficial information about these factors. Yet due to the relatively straightforward and descriptive nature of this finding, we did not deem the data thinness to be serious enough to significantly lower our confidence in the review finding. We thus concluded that we had only minor concerns about data adequacy for this review finding (Table 1). Therefore, and in summary, simpler findings may be adequately supported with less evidence and less rich evidence.

Making the adequacy assessments for the more interpretive findings was simplified considerably by the fact that the primary sampling criteria we used for our inclusion of studies in the analysis was ‘conceptual richness’. Due to this criterion, many of the included studies were situated within sociological and anthropological research traditions, where ‘thick’ descriptions of intentions, meanings and interactions are arguably more common than in public health research (Green & Thorogood, 2004). Many studies were also described across multiple sources (the 27 sampled studies were reported in a total of 53 full texts, including three books) and often published in social science journals which are frequently less stringent with word limits compared to biomedical and public health journals. For these reasons, the evidence contributing to our more interpretive findings was, in most cases, of considerable depth, detail, and breadth. Had we not used ‘conceptual richness’ as our primary sampling criterion, the threats would most likely have been bigger and the judgements harder for the more interpretative findings.

### *Component 2: Relevance*

As with the adequacy component, we assessed the relevance component by asking the same question for all our review finding types. In this case, we were interested in the extent to which the body of data from the primary studies supporting a review finding reflects or aligns with the context specified in the review question (Noyes, et al., 2018c). Again, we approached our assessments in a similar way for all our review finding types—we extracted key contextual data from the primary studies and then identified similarities and differences between the contexts of the studies supporting each review finding and the context specified in the review question. Using routine procedures when applying GRADE-CERQual, if we deemed there to be significant threats, we either lowered our confidence or reformulated the finding so as to strengthen our confidence in its relevance.

For example, for both the more descriptive and more interpretive findings in Tables 1 and 2 respectively, we had initially framed the ‘summary of findings’ without incorporating any reference to context. However, in both cases the evidence suggested that the finding may be more applicable to parents from specific economic contexts—those from lower income settings in the case of the more descriptive and those from higher income settings in the case of the interpretive finding. We deemed it more meaningful and useful to end users to rephrase both findings to indicate that they were formulated in reference to a particular economic ‘subgroup’. In line with current GRADE-CERQual guidance (Noyes, et al., 2018c), for the more descriptive finding (Table 1) we therefore added the phrase “*parents living in resource-limited settings*” and for the more interpretive finding (Table 2), we incorporated the term “*many parents, predominantly from HICs*”. In light of these revisions, in both cases we reassessed the contributing evidence as “directly relevant” to the ‘subgroup’ contexts specified in the review finding (Noyes, et al., 2018c). We therefore concluded that we had only ‘minor concerns’ and ‘no or very minor concerns’ about data relevance for the more descriptive and interpretive review findings respectively.

Once again, the main difference we experienced with our different types of findings was that, as they became more interpretive, the demands placed on the evidence increased. That is, and in the case of relevance, our confidence in the broad claims and complex associations being made in our more interpretive findings necessitated that the body of contributing studies be contextually diverse, including a range of times, places, phenomena of interest and perspectives. As with our adequacy assessments, our evaluations of the relevance component for the more interpretive findings were simplified and the potential threats reduced by the sampling approach we had employed. Our primary sampling criteria of ‘conceptual richness’ and associated inclusion of sociological and anthropological research meant that the supporting data were rich in contextual detail. As is common with these disciplines, many studies provided in-depth and nuanced descriptions of the populations, settings and perspectives, as well as the broader socio-political and historical contexts in which the research was conducted. As such, insufficient clarity or reporting of contextual details, a common threat to relevance (‘unclear relevance’), was rarely an issue. At the same time, our second sampling criteria of ‘geographical spread’ meant that the studies included in our analysis came from a range of settings, including different WHO regions, urban and rural locations, as well as high-, middle-, and low-income countries. Consequently, the threat of contributing studies only representing a subset of the review scope (‘partial relevance’) was often not present. In other words, the sampling approach we had employed meant that the evidence contributing to our more interpretive findings was, in most cases, contextually rich and diverse. This ultimately

simplified our judgements and lessened our concerns about the relevance component.

### Component 3: Methodological Limitations Component

As with the previous two components, we assessed the methodological limitations component by asking the same question for the more descriptive and more interpretive findings: to what extent do we have concerns about the design or conduct of the primary studies that contributed evidence to an individual review finding (Munthe-Kaas et al., 2018). For all our review findings, we employed an adapted version of the Critical Appraisal Skills Programme (CASP) tool (CASP, 2018) to appraise the quality of the studies, and then used these appraisals to assess whether we had any concerns regarding the methodological limitations of the body of data supporting the review finding.

When assessing methodological limitations in the context of GRADE-CERQual, the goal is not to judge whether some absolute standard of methodological quality has been achieved, but rather to indicate concerns that are serious enough to lower our confidence in relation to each specific review finding (Munthe-Kaas et al., 2018). For our review, it made intuitive sense to us that methodologically ‘weak’ studies are likely to pose more serious concerns for complex, interpretive review findings compared to simpler and more descriptive review findings. We also experienced this more concretely when considering some of the specific components of our adapted version of the CASP tool.

For example, when examining the criterion ‘*Was the data analysis described and was this appropriate?*’ we noted that many studies lacked details about the analysis process, and few interrogated the credibility of their findings through methods such as triangulation or considering evidence both for and against the arguments being made. When making our GRADE-CERQual assessments for a relatively simple, descriptive finding such as ‘Socio-economic challenges in accessing vaccination services’ (Table 1), we did not deem this absence to be serious enough to lower our confidence. We judged it unlikely that the data contributing to this finding—which essentially lists barriers to accessing vaccination services—would have been significantly different had the study authors considered potential contradictory data or conducted other sorts of credibility checks. This is because, while there is inevitably some level of interpretation embedded in all types of finding, more descriptive findings are essentially naming or categorising phenomena. There is arguably less that can go awry in the development of such findings, and in turn, less potentially required for showing the claims are sufficiently trustworthy. As such, despite many of the studies which contributed data to our review finding on ‘Socio-economic challenges in accessing vaccination services’ lacking details on the credibility of this data and how it was produced, we concluded that we had only minor concerns about methodological limitations for this review finding.



However, we deemed this same limitation to be more serious for a more complex interpretive review finding like our ‘neoliberal logic’ finding (Table 2). In this case, we considered it possible that the data supporting this finding could have been different had the study authors considered refutational interpretations or performed other forms of critical engagement with the data. For example, the studies contributing to this review finding all showed that parents’ vaccine narratives were saturated with discourses of personal responsibility, choice and individualised risk. However, the data from a few of these studies also revealed a slightly more complex picture, with some parents reflecting multiple and at times conflicting values: personal choice but also collective responsibility; the individual as the expert but also ‘doctor knows best’. Had the study authors taken these nuances further in their analyses and incorporated them more in their interpretations, they might have provided a slightly different argument about the nature and drivers of vaccine hesitancy. And as a result, we might have constructed our ‘neoliberal logic’ review finding differently, for example, by making other sorts of inferences, by adding additional nuances or qualifications, or by drawing on an alternative overarching theory through which to explain the underlying study data.

The point is that more interpretive findings are often attempting to make complex conceptual arguments about the nature and workings of phenomena. And to do this they are usually constructed out of multiple underlying claims about relationships and processes that can be inter alia descriptive, theoretical, and/or inferred. There are, therefore, many avenues through which the construction of these findings could go awry, and in turn more is arguably required to show that the claims being made are sufficiently trustworthy. We thus concluded that we had moderate concerns about methodological limitations for our ‘neoliberal logic’ finding due to insufficient evidence regarding how the data was produced and its credibility.

We recognise, however, that this link between the type of review finding and the demands placed on the methodological rigour of the contributing studies may not be straightforward or inevitable. Methodological limitations in the context of GRADE-CERQual are not absolute, but always depend on the specific review topic, the specific study, the specific finding (its content and its structure) and the specific weakness. Some methodological weaknesses may therefore be important for some reviews and review findings but not others, and the same methodological quality issues may raise different levels of concern for different review findings (Munthe-Kaas et al., 2018). Therefore, the nature of the review finding—and the extent of its interpretive complexity—is arguably one potential factor, amongst others, that needs to be considered when making the methodological limitations GRADE-CERQual assessments.

In the case of methodological limitations, and in direct contrast to the adequacy and relevance components, our assessments were made more difficult and the threats potentially

amplified by the sampling approach we had employed. Our ‘conceptual richness’ sampling criterion, and associated inclusion of sociological and anthropological research, may have contributed to the inclusion of many studies which poorly reported the methods used. Within these disciplines, there has traditionally been little emphasis on describing the processes of data collection and analysis (Green & Thorogood, 2004). Indeed, three of the studies included in our review, which made the most significant contributions to the review findings, were books with little (if any) information about methods. It was therefore often challenging to ascertain the methodological quality of the studies, and the potential impact of this on our confidence in the more interpretive review findings. Had we used an alternative appraisal tool, potentially more aligned with the methods and epistemologies of ethnographic research, the assessment process may have been easier and the recurring threat of uncertainty due to poor reporting reduced.

#### Component 4: Coherence

As with the other GRADE-CERQual components, the core principles we used and the manner in which we assessed coherence was similar for all our review finding types. In this case, we asked the same broad question: is the fit between the underlying data from the primary studies and the review finding clear and cogent? (Colvin et al., 2018). And we approached this in the same way for all our review findings - collating the underlying data from the primary studies contributing to each review finding and then assessing whether we had any concerns about the fit between the body of contributing data and the review finding. In the case of both the more descriptive and more interpretive findings, where significant threats were identified, we either lowered our confidence or reformulated the finding so as to strengthen our confidence in its coherence. For example, we had initially constructed our more descriptive finding on ‘socio-economic challenges in accessing vaccination services’ (Table 1) as follows:

‘Socio-economic challenges in accessing vaccination services’: Parents living in resource-limited settings frequently face numerous socioeconomic challenges to accessing vaccination services which reduces their acceptance of vaccination”.

On reviewing the body of evidence contributing to this finding, we found that the underlying data was in fact more varied than captured in this finding, including data that did not fit with the pattern described (‘contradictory data’). For example, in some studies there were parents who faced socio-economic challenges to accessing vaccination services yet still accepted vaccination or even went to great lengths to overcome these barriers to obtain vaccination for their children. We therefore decided that this review finding was somewhat of an over-simplified description of the patterns in the underlying data. As is common practice in the framing of any qualitative

interpretation, we therefore modified its formulation to strengthen the fit between the review finding and the data. Specifically, we rephrased the finding slightly to indicate that *“some of these parents were less accepting of vaccination”* due to these challenges (Table 1). This minor, but important qualification, ultimately improved the precision of the review finding and avoided over-generalizing.

In a similar way, we had initially framed our interpretive finding on ‘neoliberal logic’ (Table 2) with the following declarative statements:

‘Neoliberal logic’: Many parents...held a neoliberal worldview. This view understands health as... Vaccination promotion messages are underpinned by contradictory discourses, ones which emphasise...This incompatibility between vaccination promotion messages and a neoliberal worldview led parents to be less accepting of vaccination.

In a similar way to the more descriptive finding, on reviewing the body of evidence contributing to this finding we found that there was data which challenged our explanation and also data that suggested that the issues were more complex (‘contradictory data’). For example, there were many parents who held neoliberal views who accepted vaccination and who did not see vaccination programmes as incompatible for them. The underlying data from the studies also suggested that referring to a ‘neoliberal worldview’ oversimplified what were worldviews frequently made-up of a variety of discourses, including (but not limited to) neoliberal discourses. As with the more descriptive finding above, we therefore deemed the finding, as initially expressed, to have serious threats to coherence. To increase our confidence in the coherence of this review finding, we therefore modified it to better capture the nuances in the data as well as the data that challenged our interpretation. As depicted in Table 2, rather than saying: *“many parents’...held a neoliberal worldview”* we spoke about *“a worldview informed by neoliberal discourses”*. Rather than talking about the *“incompatibility between vaccination promotion messages and a neoliberal worldview”* we changed the focus to *“parents’ perceived tension”* and the fact that *“some parents experienced this worldview as in conflict with vaccination promotion messages”*. And rather than stating that these factors *“led parents to be less accepting of vaccination”* we explained that these factors *“led some parents to resist these messages and be less accepting of vaccination”*. Again, this way of making qualitative review findings more conceptually precise is usual when developing and refining any interpretation and is a common practice when applying GRADE-CERQual.

The general process we followed to assess coherence for our more descriptive and more interpretive review findings was therefore very similar. That said, the coherence assessments- more than with the other GRADE-CERQual components-generated various challenges, dilemmas and more complex judgements for the more interpretive review findings.

A first challenge was ascertaining exactly what data contributed to the more interpretive review findings. In order to make the coherence assessments, one needs a clear sense of the underlying data from the primary studies relevant to the review finding. For this reason, it is recommended that one keeps a clear and transparent ‘audit trail’ for the analysis so one can track what data contributed to each review finding (Flemming & Noyes, 2021). However, with more interpretive (as opposed to aggregative) synthesis methodologies and associated outputs, it is arguably more difficult to keep such an audit trail (Noyes, Booth, Flemming, et al., 2018). Here the findings often shift and evolve iteratively through the synthesis process in ways that “cannot be reduced to mechanistic tasks” (Britten et al., 2002). It can therefore be challenging to decipher when and why transformations in findings occur, and what specific data contributed to them.

As part of our data extraction processes, we drew heavily on the eMERGe guidance (France, Uny, et al., 2019). This guidance aims to improve the reporting of meta-ethnographies by providing detailed reporting steps and processes for each of the analysis stages commonly employed with a meta-ethnographic synthesis approach. Using this guidance proved helpful for keeping at least better track of what data contributed to the more interpretive findings and their evolution. However, when it came assessing the coherence of these findings, we frequently needed to return to the primary studies and even at times develop further coding. This was because the details necessary to make these assessments were not always captured in our original data extraction processes.

More than this, however, with more interpretive synthesis methodologies there are parts of the analysis process for which an ‘audit trail’ arguably does not actually exist. Particularly when developing more interpretive findings, one is often in more abstract territory where ‘inference’ forms a central part of the analysis. For example, with our ‘neoliberal’ review finding (Table 2) we claim to have identified a worldview and argue that something is at work with this worldview and vaccination. Yet none of the study participants spoke about neoliberalism - they reported issues such as choice, responsibility, risks and so forth and we inferred this to be an expression of a pre-existing conceptual framework for something termed ‘neoliberalism’. Similarly, none of the study authors aimed to identify, nor focused on, the worldviews of participants. As such, we essentially read our interpretation of a neoliberal logic into the words of the study participants and authors. In contrast, with the more descriptive findings, both the participants and study authors often explicitly used, or at least one could imagine them using, the words of the finding to describe or explain the phenomenon. For example, with our finding about ‘socio-economic challenges in accessing vaccination services’ (Table 1) both study participants and authors explicitly named access challenges and themselves directly attributed these to reduced vaccination acceptance.

The point is that more interpretive findings are transformed findings and are thus by their very definition less directly linkable to the data in the primary studies. The ‘fit’ between the finding and the underlying data is therefore inevitably weakened. This is indeed a defining characteristic of a meta-ethnographic synthesis approach, where the objective is to offer novel interpretations that ‘go beyond’ the data of the studies (Campbell et al., 2011; Noblit & Hare, 1988). Thus, a second dilemma we faced was should we, and if so how, incorporate the inherent threats to coherence of our more interpretive findings?

As a review team we agreed that there needs to be a way of applying the GRADE-CERQual principles where the inherent ‘distance’ of more interpretive findings can be factored into the assessment. We considered this important as, a failure to do so would mean that more interpretive findings would always be ranked as low confidence, and the coherence assessment essentially becomes a way of ranking the degree of transformation of review findings. Yet we were unsure how this should be done. An option, amongst other potential possibilities, could be to start off with the assumption that there are inherent threats to coherence with more interpretive findings. This deviates from current guidance on making coherence assessments which stipulates, like with all the GRADE-CERQual components, that we begin with the assumption that there are no concerns with coherence (Colvin et al., 2018). Beginning with this alternative assumption, we might then have criteria that could be used to ‘increase’ our confidence in coherence, perhaps comparable to the way GRADE for effectiveness reviews has criteria for ‘grading-up’ observational studies (Guyatt, Oxman, Sultan, et al., 2011). However, a difficulty with this would be how one defines a threshold of ‘interpretation’ or ‘transformation’ that allows one to flip the approach, assuming flipping the approach would be appropriate. Ultimately, further thought and discussion in this regard would be helpful.

A third, related challenge we faced when making the coherence assessments for the more interpretive findings was should we, and if so how, incorporate the multiple forms of evidence commonly forming part of the construction of these types of findings? More interpretive findings are usually developed out of the combination of various sources of evidence - theory (imported by the reviewers, identified and/or developed in the included studies and/or originally developed by the reviewers), expert opinion, reflexivity, personal experience, imagination, creativity, inference - along with the empirical data from studies. Arguably, it is impossible to develop any interpretations without some reference to pre-existing terms, categories, frameworks or theories about the world. Even personal experience and expert opinion shape what seems thinkable and possible as an explanation or interpretation. The point here is that no interpretation, especially ones rooted in highly transformed data, can emerge solely from the underlying evidence collected in a study or review.

And yet, currently the GRADE-CERQual coherence component focuses primarily on the fit between the review finding and the empirical data from the studies, although current guidance has, to some degree, included theory as a possible evidence source (Colvin et al., 2018). Drawing on this guidance, we quite substantially incorporated theory into our coherence assessments for our more interpretive findings. For example, for our ‘neoliberal logic’ finding (Table 2), we ‘imported’ from the literature, external to the studies included in the synthesis, the overarching theory of neoliberalism. We used this theory to explain the underlying empirical data and to bring together the various concepts used in the studies. In our coherence assessment of this finding, we argued that neoliberalism is a relatively well-established and developed social theory, and as such this enhances our confidence, at least to some degree, in the coherence of our finding.

However, besides theory, we wondered how the other sources of evidence that commonly support the construction of findings, particularly more interpretive ones, could be brought into the assessment of coherence. In other words, how might judgements about the credibility of review findings be broadened to incorporate more diverse evidentiary sources beyond empirical data and theoretical insights? Here it could be helpful to draw on some of the thinking that has emerged within critical social science scholarship, particularly the field of Science and Technology Studies (STS). Scholars working in this field (Bowker & Star, 1999; Elgin, 2004; Green, 2009; Haraway, 1999; Latour, 2010; Stengers, 2012; Turnbull, 2000) have for some time now demonstrated the limitations of evidence-based medicine (EBM) and its empiricist underpinnings. They have highlighted how within EBM only those aspects of ‘reality’ which are directly observable and currently measurable as empirical pieces of data are considered valid forms of evidence. Consequently, other knowledges and ways of knowing are inevitably delegitimised and in turn ignored. These ‘alternative’ sorts of evidentiaries are frequently more tacit and experiential, more emotional and embodied, more contingent and relational, and most certainly do not easily fit with the familiar kinds of abstractions of EBM. Yet, according to these scholars, these alternatives offer potentially important ways in which aspects of the social world might be constituted and articulated through. The evidence base of peer-reviewed research literature is inevitably incomplete and biased, with the attendant risk that interpretations of the world rooted solely in this literature may be potentially misleading and/or overlook crucial dimensions of perspective, experience, relationship and practice. From this perspective, it is important to complement empirical data from scientific research with other types and sources of evidence (themselves also inevitably partial and biased).

For many STS scholars then, there is a need for EBM to be more inclusive of a wider range of knowledge practices and sources and ultimately more “hospitable” to different iterations of reason and the reasonable (Green, 2009). Importantly, the argument they are making is not one of relativism, a kind

of ‘anything goes’. Nor are these scholars denying the importance of choice, judgment and critical assessment. What they are arguing for is the need to rethink our forms of judgement about evidence in ways that do not straightforwardly disqualify nor valorise whatever does not fit the epistemological canon of EBM. Ultimately, they are asking how we might work credibly, critically and more hospitably with diverse knowledges and ways of knowing.

In grappling with this, these scholars have developed various conceptual resources for potential new understandings and imaginings of scholarly acceptability. These include, for example, Stengers’ (Stengers, 2012) concept of “reclaiming animism”, Green’s (Green, 2009) “reflective equilibrium”, Turnbull’s (Turnbull, 2000) “knowledge motley”, Elgin’s (Elgin, 2004) “felicitous falsehoods” and Latour (Latour, 2010) notion of the “factish”. These challenging - yet enticing - concepts might offer potential avenues for careful and critical thinking about how more diverse evidentiary sources might be incorporated within EBM and GRADE-CERQual’s assessments of coherence.

A final dilemma we faced when making the coherence assessments for our more interpretive findings was whether we should be concerned only with refutational interpretations, or if *any* alternative interpretation(s) might be a cause for concern? As described in current guidance (Colvin et al., 2018), one of the three types of threats to coherence is ‘plausible alternatives’, which is concerned with whether there are alternative plausible ways of describing, interpreting or explaining the data and which have not been examined by the review authors. When assessing this threat for our interpretive findings, we found that there were various possible alternative interpretations and many equally valid theories that we could have been used to explain the patterns in the data. For example, for our ‘neoliberal’ finding (Table 2) we could have drawn on various theories from social psychology related to, for example, risk beliefs, appraisal and utility calculation, or social identity theory. The point is that with more interpretive findings, there are always different ways of thinking about or explaining the problem. Yet it is arguably inconceivable for the role of the GRADE-CERQual component of coherence to be about assessing the framing of the review finding against all other possible framings. As such, and what we decided upon for our assessments, was that we should be concerned only with alternative theories or explanations that specifically refute or contradict our interpretation. This does, however, raise questions around the terms by which a theory or paradigm should be considered ‘refutational’- such criteria could be epistemological, ontological, political, moral and so forth. Again, more thought and discussion on these issues would be helpful.

## Conclusions

In this paper we have reflected on our experiences of applying GRADE-CERQual to the findings that emerged from a

Cochrane meta-ethnography on childhood vaccination acceptance. Specifically, we focused on the similarities as well as the differences, challenges and dilemmas we experienced when applying the approach to more interpretive findings compared to more descriptive findings. We found that we were able to employ the core criteria and principles of GRADE-CERQual in ways that were congruent with the methodologies and epistemologies of a meta-ethnographic approach and associated more interpretive outputs. We also found that the practical application processes were similar across the spectrum of review finding types.

The main differences we found were the level of demand placed on the evidence supporting the finding and the level of complexity involved in the judgements, most particularly for the GRADE-CERQual component of coherence. With the more interpretive findings, it was more difficult for us to have the same degree of confidence in them. This was not because any criteria changed, or were applied differently, but because the same criteria and application process faced a more daunting challenge. Ultimately, the complex and often abstract nature of our more interpretive findings meant that for us to have a similarly high level of confidence in them as with our simpler, more descriptive findings, more from the supporting data was required. At the same time, the more interpretive findings involved considerably more complex forms of judgement and perhaps greater anxiety for us as review authors. Both the development of ‘summary of findings’ and the confidence assessments for the interpretive findings were more challenging, required more time, critical thought and discussion as a review team, and necessitated a particularly deep and nuanced grasp of the logic of GRADE-CERQual.

The level of complexity involved in these processes, and the concerns we faced, were heavily influenced by the sampling approach we used for our review. That is, our primary sampling criteria- ‘conceptual richness’ and ‘geographical spread’- led to a body of evidence supporting the more interpretive findings that was, for the most part, considerably rich, thick and contextually situated. In the case of the GRADE-CERQual components of adequacy and relevance, this lessened the threats and simplified our assessments. Yet our sampling approach also contributed, arguably, to the inclusion of many studies with poor reporting of methods. In the case of the component of methodological limitation, this increased the threats and further complicated our assessments for the more interpretive findings. Therefore, and as suggested elsewhere (Ames et al., 2019), when review authors develop their sampling strategy it could be helpful to consider the implications it may have on the subsequent GRADE-CERQual process. This would be beneficial for all types of review methodologies, but particularly for those of a more interpretive nature.

That said, sampling to shape the type of evidence included in the review, and associated GRADE-CERQual facilitators and challenges, is not always an available option. You also need a topic that has a large volume of research and that



includes studies of sufficient breadth and depth. For many topics, this is not the case, and you either choose not to sample because you have too few studies or you sample but are still left with conceptually thin studies or studies from very few settings. We were fortunate for our review in that the topic of childhood vaccination has been extensively studied and thus we had access to a wealth of rich studies from many different contexts (145 studies met our inclusion criteria and we sampled 27 of these for our analysis). We therefore had the opportunity to consider the type of evidence we would like to include in our review and how we might sample accordingly.

The evaluations for the more interpretive review findings were generally more complicated across the four GRADE-CERQual components, but most particularly for the coherence component. Here we faced a series of challenges and quandaries, including clearly ascertaining the underlying contributing data, questions around the significance of refutational versus alternative interpretations, and whether (and if so how) to incorporate the inherent threats of ‘distance’ and the multiple sources of evidence constituting the construction of more interpretive findings. In flagging these issues, our uncertainties surrounding them, and in some instances making preliminary suggestions for how they might be addressed, we hope to open them up for further scrutiny and debate. Such engagement could enhance the usability of GRADE-CERQual for more interpretive review findings, and in turn the potential use of these types of findings within health and social care policy- and decision-making.

Most certainly, we recognise the apprehensions within more critical qualitative research communities about the use of qualitative research within decision-making (Lambert et al., 2006; Mykhalovskiy & Weir, 2004; Sandelowski et al., 1997; Thorne et al., 2004). The concern is that exposing qualitative research to the highly technical principles and procedures of evidence-based medicine (EBM) threatens to compromise the politics and epistemologies of such research (Colvin, 2015). In other words, qualitative research that seeks to challenge dominant systems and logics and promote deep and nuanced understandings, risk being depoliticised or diminished by EBM and its positivist ideals of empiricism, rationalism, objectivity, and standardization (Timmermans & Berg, 2003). We share these trepidations, yet at the same time believe in the transformative potential of strategies- inevitably precarious- that seek to enlarge the kinds of qualitative knowledge that might contribute to decision-making processes. We see the increased use of, and potential enhancement of guidance around GRADE-CERQual for more interpretive synthesis methodologies and outputs as one such strategy. That is, it provides a way of bringing to the table rich insights and theoretical frameworks of experience and context that could potentially unsettle and expand simplistic or one-dimensional concepts that often dominate decision-making interactions (Brookfield et al., 2019). It affords a possible mechanism for broadening the kinds of issues qualitative research is typically sought for- such as ‘acceptability’ and ‘feasibility’- to more

critical conversations about ‘power’, ‘ideology’, ‘structure’ and ‘justice’ (Colvin, 2015). Ultimately, it offers potential openings and opportunities for expanding the kinds of knowledges and ways of knowing that count within health and social care decision-making.

### Acknowledgments

We would like to thank Claire Glenton who was the editor of the original review and who provided significant comments to improve the review, including the GRADE-CERQual assessments. She also provided feedback on a draft version of this paper which helped strengthen it. We would also like to thank EPOC Norway as part of the Norwegian Institute of Public Health for organising a webinar to discuss the contents described in this paper.

### 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: This work was supported by the South African Medical Research Council, the University of the Western Cape, South Africa and the Research, Evidence and Development Initiative (READ-It) (Project number 300342-104), Commonwealth and Development Office, UK. This paper has been funded by the South African Medical Research Council.

### ORCID iDs

Sara Cooper  <https://orcid.org/0000-0001-9894-236X>

Charles S. Wiysonge  <https://orcid.org/0000-0002-1273-4779>

Christopher J. Colvin  <https://orcid.org/0000-0002-8930-7863>

### References

- Ames, H., Glenton, C., & Lewin, S. (2019). Purposive sampling in a qualitative evidence synthesis: A worked example from a synthesis on parental perceptions of vaccination communication. *BMC Medical Research Methodology*, 19(1), 26. <https://doi.org/10.1186/s12874-019-0665-4>
- Bohren, M., Glenton, C., Noyes, J., Wainwright, M., & Lewin, S. (2023). Chapter 13. Assessing confidence in the evidence using the GRADE-CERQual approach. Draft version (February 2024) for inclusion in: J Noyes, & A Harden, editor(s). *Cochrane-Campbell Handbook for Qualitative Evidence Synthesis*, Version 1. Cochrane.
- Bowker, G., & Star, S. (1999). *Sorting things out: Classification and its consequences*. MIT Press.
- Britten, N., Campbell, R., Pope, C., Donovan, J., Morgan, M., & Pill, R. (2002). Using meta ethnography to synthesise qualitative research: A worked example. *Journal of Health Services Research & Policy*, 7(4), 209–215. <https://doi.org/10.1258/135581902320432732>
- Brookfield, S., Fitzgerald, L., Selvey, L., & Maher, L. (2019). The blind men and the elephant: Meta-ethnography 30 Years on.

- Qualitative Health Research*, 29(11), 1674–1681. <https://doi.org/10.1177/1049732319826061>
- Campbell, R., Pound, P., Morgan, M., Daker-White, G., Britten, N., Pill, R., & Donovan, J. (2011). Evaluating meta-ethnography: Systematic analysis and synthesis of qualitative research. *Health Technology Assessment*, 15(43), 1–164. <https://doi.org/10.3310/hta15430>
- Carmona, C., Baxter, S., & Carroll, C. (2021). Systematic review of the methodological literature for integrating qualitative evidence syntheses into health guideline development. *Research Synthesis Methods*, 12(4), 491–505. <https://doi.org/10.1002/jrsm.1483>
- CASP. (2018). *Critical appraisal Skills programme (CASP) qualitative checklist: 10 questions to help you make sense of a qualitative research*. CASP. Available at: <https://casp-uk.net/wp-content/uploads/2018/01/CASP-Qualitative-Checklist.pdf>
- Colvin, C. J. (2015). Anthropologies in and of evidence making in global health research and policy. *Medical Anthropology*, 34(2), 99–105. <https://doi.org/10.1080/01459740.2014.963196>
- Colvin, C. J., Garside, R., Wainwright, M., Munthe-Kaas, H., Glenton, C., Bohren, M. A., & Lewin, S. (2018). Applying GRADE-CERQual to qualitative evidence synthesis findings-paper 4: How to assess coherence. *Implementation Science*, 13(Suppl 1), 13. <https://doi.org/10.1186/s13012-017-0691-8>
- Cooper, S., Schmidt, B. M., Sambala, E. Z., Swartz, A., Colvin, C. J., Leon, N., & Wiysonge, C. S. (2021). Factors that influence parents' and informal caregivers' views and practices regarding routine childhood vaccination: A qualitative evidence synthesis. *Cochrane Database of Systematic Reviews*, 10(10), Cd013265. <https://doi.org/10.1002/14651858.CD013265.pub2>
- Downe, S., Finlayson, K. W., Lawrie, T. A., Lewin, S. A., Glenton, C., Rosenbaum, S., & Tunçalp, O. (2019). Qualitative Evidence Synthesis (QES) for Guidelines: Paper 1 - using qualitative evidence synthesis to inform guideline scope and develop qualitative findings statements. *Health Res Policy Syst*, 17(1), 76. <https://doi.org/10.1186/s12961-019-0467-5>
- Elgin, C. (2004). True enough. In E. S. Villanueva (Ed.), *Epistemology: Philosophical issues* (pp. 113–131). Blackwell Publishing.
- Fleming, K., & Noyes, J. (2021). Qualitative evidence synthesis: Where are we at? *International Journal of Qualitative Methods*, 20(January-December), 1609406921993276. <https://doi.org/10.1177/1609406921993276>
- France, E. F., Cunningham, M., Ring, N., Uny, I., Duncan, E. A. S., Jepson, R. G., & Noyes, J. (2019). Improving reporting of meta-ethnography: The eMERGe reporting guidance. *BMC Medical Research Methodology*, 19(1), 25. <https://doi.org/10.1186/s12874-018-0600-0>
- France, E. F., Uny, I., Ring, N., Turley, R. L., Maxwell, M., Duncan, E. A. S., & Noyes, J. (2019). A methodological systematic review of meta-ethnography conduct to articulate the complex analytical phases. *BMC Medical Research Methodology*, 19(1), 35. <https://doi.org/10.1186/s12874-019-0670-7>
- Glenton, C., Carlsen, B., Lewin, S., Munthe-Kaas, H., Colvin, C. J., Tunçalp, Ö., & Wainwright, M. (2018). Applying GRADE-CERQual to qualitative evidence synthesis findings-paper 5: How to assess adequacy of data. *Implementation Science*, 13(Suppl 1), 14. <https://doi.org/10.1186/s13012-017-0692-7>
- Glenton, C., Lewin, S., Lawrie, T. A., Barreix, M., Downe, S., Finlayson, K. W., & Tunçalp, Ö. (2019). Qualitative Evidence Synthesis (QES) for Guidelines: Paper 3 - using qualitative evidence syntheses to develop implementation considerations and inform implementation processes. *Health Res Policy Syst*, 17(1), 74. <https://doi.org/10.1186/s12961-019-0450-1>
- Green, G., & Thorogood, N. (2004). *Qualitative methods for health research*. Sage.
- Green, L. (2009). Challenging epistemologies: Exploring knowledge practices in Palikur astronomy. *Futures*, 41(1), 41–52.
- Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., & Schünemann, H. J. (2011). GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*, 64(4), 383–394. <https://doi.org/10.1016/j.jclinepi.2010.04.026>
- Guyatt, G., Oxman, A. D., Sultan, S., Glasziou, P., Akl, E. A., Alonso-Coello, P., & Schünemann, H. J. (2011). GRADE guidelines: 9. Rating up the quality of evidence. *Journal of Clinical Epidemiology*, 64(12), 1311–1316. <https://doi.org/10.1016/j.jclinepi.2011.06.004>
- Hannes, K., & Macaitis, K. (2012). A move to more systematic and transparent approaches in qualitative evidence synthesis: Update on a review of published papers. *Qualitative Research*, 12(4), 402–442. <https://doi.org/10.1177/1468794111432992>
- Haraway, D. (1999). Situated knowledges: The science question in feminism and the privilege of partial perspective. In M. Biagioli (Ed.), *The science study reader* (pp. 172–188). Routledge.
- Hsu, J., Brozek, J. L., Terracciano, L., Kreis, J., Compalati, E., Stein, A. T., & Schünemann, H. J. (2011). Application of GRADE: Making evidence-based recommendations about diagnostic tests in clinical practice guidelines. *Implementation Science*, 6, 62. <https://doi.org/10.1186/1748-5908-6-62>
- Lambert, H., Gordon, E. J., & Bogdan-Lovis, E. A. (2006). Introduction: Gift horse or Trojan horse? Social science perspectives on evidence-based health care. *Social Science & Medicine*, 62(11), 2613–2620. <https://doi.org/10.1016/j.socscimed.2005.11.022>
- Langlois, E. V., Tunçalp, Ö., Norris, S. L., Askew, I., & Ghaffar, A. (2018). Qualitative evidence to improve guidelines and health decision-making. *Bulletin of the World Health Organization*, 96(2), 79–79a. <https://doi.org/10.2471/blt.17.206540>
- Latour, B. (2010). *On the modern cult of the factish gods*. Duke University Press.
- Lewin, S., Bohren, M., Rashidian, A., Munthe-Kaas, H., Glenton, C., Colvin, C. J., & Carlsen, B. (2018). 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. *Implementation Science*, 13(Suppl 1), 10. <https://doi.org/10.1186/s13012-017-0689-2>
- Lewin, S., Booth, A., Glenton, C., Munthe-Kaas, H., Rashidian, A., Wainwright, M., & Noyes, J. (2018). Applying GRADE-

- CERQual to qualitative evidence synthesis findings: Introduction to the series. *Implementation Science*, 13(Suppl 1), 2. <https://doi.org/10.1186/s13012-017-0688-3>
- Lewin, S., & Glenton, C. (2018). Are we entering a new era for qualitative research? Using qualitative evidence to support guidance and guideline development by the world health organization. *International Journal for Equity in Health*, 17(1), 126. <https://doi.org/10.1186/s12939-018-0841-x>
- Lewin, S., Glenton, C., Lawrie, T. A., Downe, S., Finlayson, K. W., Rosenbaum, S., & Tunçalp, O. (2019). Qualitative Evidence Synthesis (QES) for Guidelines: Paper 2 - using qualitative evidence synthesis findings to inform evidence-to-decision frameworks and recommendations. *Health Res Policy Syst*, 17(1), 75. <https://doi.org/10.1186/s12961-019-0468-4>
- Lewin, S., Glenton, C., Munthe-Kaas, H., Carlsen, B., Colvin, C. J., Gulmezoglu, M., & Rashidian, A. (2015). 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*, 12(10), Article e1001895. <https://doi.org/10.1371/journal.pmed.1001895>
- Munthe-Kaas, H., Bohren, M. A., Glenton, C., Lewin, S., Noyes, J., Tunçalp, Ö., & Carlsen, B. (2018). Applying GRADE-CERQual to qualitative evidence synthesis findings—paper 3: How to assess methodological limitations. *Implementation Science*, 13(1), 9. <https://doi.org/10.1186/s13012-017-0690-9>
- Mykhalovskiy, E., & Weir, L. (2004). The problem of evidence-based medicine: Directions for social science. *Social Science & Medicine*, 59(5), 1059–1069. <https://doi.org/10.1016/j.socscimed.2003.12.002>
- Noblit, G., & Hare, R. (1988). *Meta-ethnography: Synthesizing qualitative studies*. Sage.
- Noyes, J., Booth, A., Cargo, M., Flemming, K., Garside, R., Hannes, K., & Thomas, J. (2018a). Cochrane qualitative and implementation methods group guidance series-paper 1: Introduction. *Journal of Clinical Epidemiology*, 97, 35–38. <https://doi.org/10.1016/j.jclinepi.2017.09.025>
- Noyes, J., Booth, A., Flemming, K., Garside, R., Harden, A., Lewin, S., & Thomas, J. (2018b). 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. *Journal of Clinical Epidemiology*, 97, 49–58. <https://doi.org/10.1016/j.jclinepi.2017.06.020>
- Noyes, J., Booth, A., Lewin, S., Carlsen, B., Glenton, C., Colvin, C. J., & Munthe-Kaas, H. (2018c). Applying GRADE-CERQual to qualitative evidence synthesis findings—paper 6: How to assess relevance of the data. *Implementation Science*, 13(Suppl 1), 4. <https://doi.org/10.1186/s13012-017-0693-6>
- Sandelowski, M., & Barroso, J. (2007). *Handbook for synthesizing qualitative research*. Springer Publishing Company.
- Sandelowski, M., Docherty, S., & Emden, C. (1997). Focus on qualitative methods. *Qualitative metasynthesis: Issues and techniques*. *Research in Nursing & Health*, 20(4), 365–371. [https://doi.org/10.1002/\(sici\)1098-240x\(199708\)20:4<365::aid-nur9>3.0.co;2-e](https://doi.org/10.1002/(sici)1098-240x(199708)20:4<365::aid-nur9>3.0.co;2-e)
- Stengers, I. (2012). Reclaiming animism. *E-flux*, 36(July), 1–10.
- Thorne, S., Jensen, L., Kearney, M. H., Noblit, G., & Sandelowski, M. (2004). Qualitative metasynthesis: Reflections on methodological orientation and ideological agenda. *Qualitative Health Research*, 14(10), 1342–1365. <https://doi.org/10.1177/1049732304269888>
- Timmermans, S., & Berg, M. (2003). *The gold standard: The challenge of evidence-based medicine and standardization in health care*. Temple University Press.
- Turnbull, D. (2000). *Masons, tricksters and cartographers: Comparative studies in the sociology of scientific and indigenous knowledge*. Harwood Academic Publishers.
- Wainwright, M., Zahroh, R. I., Tunçalp, Ö., Booth, A., Bohren, M. A., Noyes, J., & Lewin, S. (2023). The use of GRADE-CERQual in qualitative evidence synthesis: An evaluation of fidelity and reporting. *Health Res Policy Syst*, 21(1), 77. <https://doi.org/10.1186/s12961-023-00999-3>