The University Of Sheffield.

Qualitative Evidence Synthesis in health and social care guidelines: what is its role and how might it be optimised?

Christopher Carmona

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield
Sheffield Centre for Health and Related Research (SCHARR)

Division of Population Health

School of Medicine and Population Health

Submitted: 18th March 2025

Revised: 25th July 2025

ACKNOWLEDGEMENTS

My thanks go firstly to my supervisors at the University of Sheffield, Dr Chris Carroll and Dr Susan Baxter, who went above and beyond the call of duty and continued to be a great supporter and sounding board even after her retirement from the University.

Also, academically, historical thanks to Dr Jo Frankham (then of UEA, Norwich) for waking the love of qualitative methods in me some three decades ago, and to Professors Mildred Blaxter and Mike Kelly who over the following two decades nurtured that love of qualitative methods and focussed it onto evidence-based health care.

Love and respect to my husband and friends who have put up with my frequent crabbiness with good grace (in the main) and managed to find the necessary balance of leaving me alone and gently encouraging me.

My thanks to NICE for allowing me access to their committees for recruiting participants, and to the staff within NICE who facilitated that access for me, and of course my eternal gratitude to the very busy health professionals, academics and NICE staff who were willing to share their experience and expertise in guideline production so freely.

ABSTRACT

Purpose and scope: The body of research reported in this thesis aims to better understand the role of qualitative evidence syntheses (QES) in the development of health (including public health) and social care guideline development, and how its role might be optimised.

Methods: Three studies were undertaken:

- A systematic review of methods and processes used for incorporating QES into health and social care guideline development.
- A quantitative content analysis to understand how frequently QES was undertaken by a leading health and social care guideline producer in the United Kingdom (the National Institute for Health and Care Excellence), and how closely the reports of those syntheses adhere to established reporting standards.
- A qualitative interview study to explore in-depth the views and perceptions of key committee participants regarding QES in health and social care, and how evidence from QES currently informs the process of developing recommendations.

Results: Key results indicate that, although there is some guidance internationally about how to undertake and integrate QES in guidelines, committee participants in this study struggle with QES, primarily in terms of understanding the purpose, role, and value of the qualitative paradigm, especially in situating it in relation to quantitative approaches. However, QES are perceived as a genuinely valuable tool for increasing stakeholder and lay-member visibility and inclusion.

Conclusions: Methods for using QES in the development of guidelines are underdeveloped and this study of one organisation in the UK indicates that further training for both committee members and staff, alongside methodological development, will enhance the usefulness of this important source of evidence. Further research is suggested alongside practical recommendations for guideline producers.

TABLE OF CONTENTS

Acknowledgements....2
Abstract....3
Table of Contents....4
List of tables....7
List of figures....8
Abbreviations....9
Declaration....11
Chapter 1: Introduction and background....12
Introduction....12
Thesis aims....14
Thesis outline....14
Reflexivity and insider researcher statement....16
Definitions....20
Background....22
The rise (and falter?) of evidence-based medicine/healthcare....22
The limitations of systematic reviews of randomised controlled trials as the mainstay of evidence-based guidelines....27
The evolution and impact of qualitative evidence synthesis in health and social care research – where are we now?....30
A critical examination of the role of QES in developing health and social care guidelines....34
Future developments....36
Summary....41
Chapter 2: Systematic review of the methodological literature for integrating qualitative evidence syntheses into guideline development....43
Introduction....43
Review question....46
Protocol....46
Methods....46
Results....49
Discussion....73
Limitations of the Review....80
Conclusions....81
Included papers....82
Chapter 3: The conduct and reporting of qualitative evidence syntheses in health and social care guidelines: A content analysis....85
Abstract....85
Introduction....86
Method....89
Results....92
Discussion....102
Conclusions....108
Chapter 4: Methods for the primary qualitative study of committee member and technical staff views and opinions on the usefulness of QES in guideline development....109
Introduction....109
Context and setting for the qualitative study....110
Research question....112

Epistemological underpinning of the qualitative study....113
Research design....114
Participants....114
Data collection methods....116
Data analysis....116
Ethics....118
Topic guide....119
Data recording and transcription....121
Ensuring trustworthiness and rigour....122
Conclusion....123
Chapter 5: Results of the primary qualitative study of committee member and technical staff views and perceptions about QES in guideline development....124
Introduction....124
Participants....124
Themes....125
Trusting the credibility and robustness of qualitative research....127
Introduction to the Theme....127
Understanding the Paradigm....127
Transparency of Reporting....130
Analytic Uncertainty....131
Generalisability....132
Presenting the findings to committees....133
Introduction to the Theme....133
Volume of Material....134
Trust in Technical Team....135
Order of Information....136
Value of qualitative findings for guideline development....137
Introduction to the Theme....137
Status of Qualitative vs. Quantitative Research....138
Lived Experience and Broader Perspectives....141
Lay-member Contributions....142
Conclusions....144
Chapter 6: Discussion of findings....145
Introduction....145
The interdependency of the themes and sub-themes....145
Central themes and their challenges....147
Trust and Transparency....147
Challenges of Presenting Large Volumes of Complex Data....148
Status of Qualitative vs. Quantitative Research....150
Interplay between quantitative and qualitative evidence....152
Implicit biases and hierarchies....153
Value of Lived Experience....155
Chapter 7: Summary and recommendations....157
Introduction....157
Optimising the use of QES in guideline development....157
Increasing Trust in the Credibility and Robustness of QES....157
Presenting Findings....159
Value....161
Limitations of this thesis....162
Methodological Limitations....162
Sampling issues....162
Measurement tools....163
Data Limitations....163

Data availability/quality 163
Data scope 164
Practical Constraints 165
Resources 165
Future research 165
Chapter 8: Conclusion 167
Chapter 9: References 170
Appendices 183
Appendix A: review protocol 184
Appendix B: systematic search history 192
Appendix C: study inclusion checklist 202
Appendix D: papers excluded at full text 207
Appendix E: Systematic review of the methodological literature for integrating qualitative evidence syntheses into health guideline development – accepted version 215
Appendix F: ENTREQ criteria, descriptions and coding rules 273
Appendix G: R code used to generate figures 277
Appendix H: Consolidated criteria for reporting qualitative studies (COREQ): 32-item checklist 279
Appendix I: Participant information sheet 284
Appendix J: Consent form 289
Appendix K: Letter of ethical approval for qualitative study 291
Appendix L: The move towards living systematic reviews and living guidelines in healthcare: consideration of the possibilities and challenges for living qualitative evidence syntheses 292

LIST OF TABLES

Table 1: Brief details of included papers 51
Table 2: Data extraction structure from NVivo - methods and processes proposed in the methodological literature for incorporating QES into evidence-based health guideline development 54
Table 3: Summary of Bengtsson method for content analysis 89
Table 4: Summary of ENTREQ criteria 91
Table 5: Prevalence of QES by guideline type 93
Table 6 Overview of participant characteristics 125
Table 7: Review protocol 184
Table 8: Summary of search results 192
Table 9: Search strategies - medline 193
Table 10: Search strategies - medline in-process 194
Table 11: Search strategies - embase 196
Table 12: Search strategies - CINAHL 198
Table 13: Search strategies - PsycINFO 199
Table 14: Search strategies - supplementary 201
Table 15: Study inclusion checklist 202
Table 16: Papers excluded at full text 207
Table 17: Coding rules for specific ENTREQ criteria 273
Table 18: COREQ checklist 279

LIST OF FIGURES

Figure 1: PRISMA flow diagram 50
Figure 2: Coding frequency 56
Figure 3: Frequency of QES by number of included papers 94
Figure 4: Number of QES published by year (2015-2019) 95
Figure 5: Number of QES meeting each ENTREQ reporting criterion (out of a total of 90) 97
Figure 6: ENTREQ criteria (out of a maximum of 21) reported by year and authoring centre 100
Figure 7: Median number of criteria in the ENTREQ framework met 101
Figure 8: Example strong and weak recommendations from a NICE guideline 112
Figure 9 Themes and sub-themes 126
Figure 10: Structure of presentation of themes 146
Figure 11: Key links between sub-themes 147
Figure 12: Hierarchy of evidence 151

ABBREVIATIONS

AI	Artificial Intelligence
ASSIA	Applied Social Science Index and Abstracts
BMJ	British Medical Journal
CAQDAS	Computer-assisted (or aided) qualitative data analysis software
CASP	Critical Appraisal Skills Programme
CERQual	Confidence in the Evidence from Reviews of Qualitative Research
CINAHL	Cumulative Index of Nursing and Allied Health Literature
COVID-19	Coronavirus disease (SARS-COV2)
DACEHTA	Danish Centre for Health Technology Assessments (HTA)
EBHC	Evidence-Based Health (& social) Care
EBM	Evidence-Based Medicine
EtD	Evidence-to-Decision
GRADE	Grading of Recommendations Assessment, Development and Evaluation
JBI	Joanna Briggs Institute
LG	Living guideline
LLM	Large Language Model
LSR	Living systematic review
NCET	Australian National Clinical Evidence Taskforce
NICE	National Institute for Health and Care Excellence (UK)
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
QES	Qualitative Evidence Synthesis/Qualitative Evidence Syntheses[1]
QIMG	Cochrane Qualitative and Implementation Methods Group
RCT	Randomised Controlled Trial
SBU	Swedish Agency for Health Technology Assessment and Assessment of Social Services

[1] In line with common usage in the literature QES has been used as an abbreviation for both the singular and plural.

SCIE	Social Care Institute for Excellence
SIGN	Scottish Intercollegiate Guidelines Network
SR	Systematic Review (with or without meta-analysis)
SSCI	Social Science Citation Index
WHO	World Health Organization

DECLARATION

I, the author, confirm that the thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not previously been presented for an award at this, or any other, university.

Three publications are associated with this thesis:

Parts of Chapter 1 summarise a paper published in Systematic Reviews in 2023.

> Carmona, C., Carroll, C. & Baxter, S. The move towards living systematic reviews and living guidelines in healthcare: consideration of the possibilities and challenges for living qualitative evidence syntheses. Syst Rev **12**, 47 (2023). https://doi.org/10.1186/s13643-023-02218-0

An abridged version of Chapter 2: 'Systematic review of the methodological literature for integrating qualitative evidence syntheses into guideline development' was published in 2021.

> Carmona C, Baxter S, Carroll C. Systematic review of the methodological literature for integrating qualitative evidence syntheses into health guideline development. Res Syn Meth. 2021; 12: 491–505. https://doi.org/10.1002/jrsm.1483

The full proof version of the paper is included in appendix E.

Chapter 3: 'The conduct and reporting of qualitative evidence syntheses in health and social care guidelines: A content analysis' is the proof version of a paper published in BMC Medical Research Methodologies in 2022.

> Carmona, C., Baxter, S. & Carroll, C. The conduct and reporting of qualitative evidence syntheses in health and social care guidelines: a content analysis. BMC Med Res Methodol **22**, 267 (2022). https://doi.org/10.1186/s12874-022-01743-1

All papers were written by me with the support of my supervisory team and published open access through the University open access fee-paying scheme and are clearly marked in the text.

CHAPTER 1: INTRODUCTION AND BACKGROUND

Introduction

This thesis investigates the role of qualitative evidence syntheses (QES) in the development of health and social care guidelines, with a particular focus on how their use can be optimised to enhance decision-making processes. Using the National Institute for Health and Care Excellence (NICE) as a case study, the research explores how QES are reported, interpreted, and used by guideline development committees. It aims to identify practical strategies to improve the integration of QES into guideline development, thereby contributing to more context-sensitive, inclusive, and experience-informed recommendations.

The research begins with a systematic review of the published literature on methods for conducting and applying QES in guideline development. While the systematic review reports that the methods suggested for using QES in guideline development are broadly consistent, it also reports that the published methods often lack detail regarding how QES should be reported to committees for their deliberation.

To address this gap, a second study examines records of how QES have been reported in NICE committee documents. A content analysis was conducted to assess the frequency with which QES are undertaken and the quality of QES reporting within NICE guidelines over a five-year period. The analysis reveals considerable variation in the reporting of QES. Teams with expertise in qualitative methods tend to report QES more comprehensively, whereas teams with a predominantly quantitative background often demonstrate weaker reporting practices. This observation underscores the need for clearer guidance and support in the integration of qualitative evidence.

While these two pieces of work permit some assessment of what should be done or has been done in terms of conduct and reporting of QES for guideline committees, a major research gap remained in terms of how guideline committees actually engage with QES. The thesis therefore includes a third study, a primary qualitative study involving committee members and technical staff at NICE. This study explores their understanding and use of QES and its perceived utility in supporting evidence-based recommendations. The findings

suggest that limited familiarity with qualitative paradigms can hinder the effective use of QES. Committees would benefit from enhanced training in pluralist research methodologies and clearer, more accessible reporting of QES findings.

The study also highlights the importance of presenting QES findings in a manner that is both comprehensive and digestible. Committees expressed a desire for presentations that allow for critical engagement without requiring blind trust in the evidence synthesis team. Furthermore, the research identifies opportunities to improve the transition from evidence to recommendations, particularly through the use of tools such as Evidence to Decision (EtD) frameworks that explicitly incorporate qualitative dimensions such as acceptability and equity.

A key insight from the study is the potential of QES to amplify the voices of populations often underrepresented in quantitative research. This can help committees to better address health inequalities and tailor recommendations for diverse lived experiences. Lay members of committees, in particular, may play a pivotal role in interpreting and championing QES findings, using them to reflect on and extend their own experiences and to represent a broader range of perspectives.

This thesis makes an original contribution to the field by focusing on the interpretation and application of QES by guideline committees—an area that has received limited attention to date. While there is growing interest in the use of QES by organisations such as WHO, most existing research has concentrated on methodological development rather than practical implementation.

The research builds on the foundational work of scholars in the Cochrane Qualitative and Implementation Methods Group and others. It extends their contributions by exploring the organisational and cultural factors that shape the use of QES in real-world settings. In doing so, the thesis provides empirical insights and practical recommendations that are already influencing NICE's approach to QES and committee preparation.

This research does not only represent the culmination of 6 years of part-time study at the University of Sheffield, but also the lessons learned from a 20-year career in health and social care guideline development with a genuine intellectual curiosity, and a desire to

balance pragmatism and methodological robustness to produce the best possible guidelines for use in the NHS and internationally. My interest in qualitative research developed as a young post-graduate student and led me, after an MSc in health science, to undertake a post-graduate certificate in qualitative research methods with the Centre for Applied Research in Education (CARE) at the University of East Anglia. In this thesis I have reflected on how and where my current and previous experiences may have influenced this research study.

Thesis aims

Overall question:

What is the role of qualitative evidence syntheses in the development of health (including public health) and social care guideline development, and how might its role be optimised?

Sub-questions:

1. What methods and processes have been developed or proposed for incorporating QES into health guideline development? (see Chapter 2)
2. How frequently were QES undertaken in the context of guideline development by an exemplar organisation in the UK over a 5-year period, and how closely do the reports of those syntheses adhere to established reporting standards? (see Chapter 3)
3. How do guideline committees at NICE use findings from QES to inform recommendation-making? (see Chapters 4-6)
4. What are the views and perceptions of technical staff, committee experts and committee lay-members regarding how a QES contributes to committee discussions, and to the process of making recommendations? (see Chapters 4-6)
5. What might be learned about best practice? (see Chapter 7)

Thesis outline

Chapter 1 contains introductory comments and descriptive background and context setting. It looks at the opportunities and challenges for QES as we move forward into a world where

artificial intelligence and large language models are becoming the norm in evidence synthesis. Chapter 1 includes a summary of a short, published paper

Carmona C, Carroll C, Baxter S. The move towards living systematic reviews and living guidelines in healthcare: consideration of the possibilities and challenges for living qualitative evidence syntheses. Syst Rev. 2023 Mar 16;12(1):47. https://doi.org/10.1186/s13643-023-02218-0

Chapter 2 sets the groundwork for the thesis by reporting the results of a systematic review of the available methodological literature to explore and summarise what is known about best practice in incorporating evidence from QES in the development of health guidelines, from development of review protocols through to reporting standards. A shortened version of the chapter was published in 2021; however, the unabridged version has been included in the thesis because it contains additional information that was removed from the journal paper to achieve word count limits. The published version can be found at

Carmona C, Baxter S, Carroll C. Systematic review of the methodological literature for integrating qualitative evidence syntheses into health guideline development. Res Syn Meth. 2021; 12: 491–505. https://doi.org/10.1002/jrsm.1483

Building on the systematic review, **Chapter 3** is the first part of a deeper dive into an exemplar guideline-producing organisation, the UK's National Institute for Health and Care Excellence (NICE). The reasons for the choice are explained in the relevant chapters. In Chapter 3, a quantitative content analysis methodology is applied to all of the QES produced by NICE over a recent 5-year period. The primary measurement is against the ENTREQ reporting criteria. These criteria were chosen because the ability of technical staff and committee members to make recommendations and interpretations of the evidence is in large part dictated by the quality and completeness of the reporting. Chapter 3 takes the form of a paper detailing the content analysis and its results that was published in BMC Medical Research Methodologies in 2022.

Carmona, C., Baxter, S. & Carroll, C. The conduct and reporting of qualitative evidence syntheses in health and social care guidelines: a content analysis. BMC Med Res Methodol **22**, 267 (2022). https://doi.org/10.1186/s12874-022-01743-1

With the contextual work of Chapters 2 & 3 complete, the following two chapters detail a primary qualitative study that aims to better understand how NICE committee members and technical staff understand QES and how useful they think it is in supporting decision-making and recommendation-making. **Chapter 4** sets out the methods for the study, and **Chapter 5** provides a detailed report of the main themes that arose from the study.

Chapter 6 takes a higher level view of the results of the qualitative study and relates the results of the study to broader themes in the field of evidence synthesis and guideline production, as well as drawing in relevant information from earlier chapters.

Finally, **Chapter 7** presents some concluding comments, including a series of practical suggestions for organisations producing health and social care guidelines in the UK and beyond to consider implementing to optimise the usefulness of QES to guideline committees. The chapter also discusses limitations of this thesis and the studies in it.

Reflexivity and insider researcher statement

Introduction to Insider Research

Insider research refers to qualitative inquiry conducted by researchers who are members of the group or organisation they are studying. In the context of health research, this often includes clinicians, public health professionals, or staff members embedded within health organisations such as the National Institute for Health and Care Excellence (NICE). Given NICE's unique role in producing evidence-based guidelines for the UK health and social care system, insider researchers can play a crucial role in illuminating the often-complex organisational processes behind policy development.

One of the principal advantages of insider research is access. Researchers who work within or closely alongside the organisation are more likely to gain access to internal meetings, documents, and decision-making processes that external researchers may find opaque or inaccessible (1). They are also well-positioned to understand the context-specific language, values, and norms that shape the organisations operations. This positionality can facilitate richer data collection and more contextually grounded analyses. Moreover, shared experiences between researchers and participants can foster trust and openness, leading to more in-depth discussions (2).

However, insider research also poses specific risks, particularly within hierarchical and politically sensitive settings like NICE. Role duality—where the researcher is simultaneously a colleague or stakeholder—may blur ethical boundaries and complicate issues of informed consent and confidentiality (3). Insiders may take certain norms or behaviours for granted, leading to uncritical assumptions or overlooked data (4). Furthermore, pre-existing assumptions and relationships may lead to confirmation bias or suppress critical scrutiny of organisational practices (Confirmation bias refers to the tendency of researchers to favour, seek out, or interpret data in ways that confirm their pre-existing beliefs, assumptions, or experiences, while potentially overlooking or undervaluing evidence that contradicts them. In insider research, this bias can be particularly pronounced due to the researcher's familiarity with the setting and participants, which may lead to uncritical acceptance of shared norms or practices) (5). These concerns are heightened in high-stakes environments such as guideline development, where research findings may have direct policy implications (6).

To mitigate these risks, insider researchers must engage in robust reflexive practice. Keeping detailed reflexive journals and participating in regular peer debriefings can help uncover and address personal biases (7). Establishing clear boundaries between professional and research roles is essential, as is securing informed consent through transparent communication about the dual role of the researcher. Additionally, triangulation, external auditing, and member checking can help ensure the credibility and trustworthiness of findings.

Insider research holds particular promise for exploring the internal workings of health policy organisations such as NICE. By leveraging their insider status responsibly, researchers can provide valuable insights into guideline development and organisational decision-making. However, careful attention to methodological rigour and ethical integrity is essential to balance the opportunities against the inherent risks.

Mitigation strategies used in the studies described in this thesis

To address the criticisms and potential pitfalls of insider research in this PhD, several strategies were employed:

Reflexivity

Engaging in reflexive practice involving continuous self-examination regarding the researcher's positionality, assumptions, and influence on the research process. This helps in maintaining critical distance and objectivity (see below for further details about reflexivity).

Peer Debriefing

Involving external peers (PhD supervisors) to challenge interpretations and assumptions and provide an additional layer of scrutiny to enhance the credibility of the research.

Transparency

Being open about one's insider status and the potential biases it may introduce in the research write-up helps in maintaining ethical standards and trustworthiness.

Introduction to Reflexivity

Reflexivity in qualitative research involves a critical self-examination of the researcher's role and potential biases that may influence the research process, for example the impact of being an 'insider' as detailed above. Reflexivity is a continual process of reflecting on one's own values, experiences, and interactions with the research context to understand how these elements may shape the research outcomes (8) and to enhance the credibility and trustworthiness of qualitative studies by making the researcher's positionality transparent (7). By engaging in reflexivity, researchers acknowledge and address their influence on the research process, thereby contributing to the overall rigour and validity of their study (9).

Personal Background and Positioning

As a technical adviser in guideline development for NICE, my interest in QES stems from both my academic background in qualitative research in health science research and from my work experience seeing committees wrestle with QES. My perspective is also shaped by being a middle-aged, white, middle-class male from a Western cultural background, and these identities may influence my interactions with both people and data.

Motivation, Preconceptions and Known Biases

My primary motivation for undertaking this PhD was to understand the ways in which QES can support the development of recommendations in evidence-based health and social care guidelines. This motivation was rooted in a longstanding belief in the value of qualitative research and its potential to capture the lived experiences, values, and contextual nuances often absent from quantitative data alone. I began this research with a strong conviction

that QES was not only a valid but also an essential method for synthesising evidence in complex, person-centred areas of health and social care.

This belief, while a source of commitment and enthusiasm, also represented a potential source of bias. I was aware that my positive orientation toward QES might have led me to overemphasise its strengths or underplay its limitations. For example, I may have been predisposed to interpret findings in a way that supported the integration of QES into guideline development, or to view resistance to QES among stakeholders as a knowledge gap rather than a legitimate methodological concern.

Additionally, my professional background and prior exposure to qualitative research had shaped my expectations about what constituted “good” or “rigorous” qualitative work. This may have influenced how I evaluated the quality of QES in the content analysis or how I interpreted participant responses in the qualitative interviews. I also recognised a potential bias toward viewing guideline development processes through a critical lens, particularly where they appeared to privilege quantitative evidence or marginalise qualitative contributions.

Throughout the research process—especially during data collection and analysis—I made a conscious effort to bracket these preconceptions. I employed strategies such as memo writing, peer debriefing, and reflexive journaling to surface and interrogate my assumptions. I also sought to remain open to findings that challenged my initial beliefs and to represent participant perspectives faithfully, even when they diverged from my own views.

Interactions With Participants

I am aware that my insider status with both committee members and with other members of technical NICE staff who participated in interviews is not neutral, and perceptions of my status as an ‘officer’ of NICE (in the case of committee members) or as a colleague or superior (in the case of NICE technical staff) might influence the ways that people speak to me about their experiences, and that there is a risk they will censor what they say, or tell me what they think is the ‘correct’ answer rather than their opinion. During all of the interviews I tried to mitigate this risk by explaining to participants as part of my introduction that I was not doing this work on behalf of NICE or with any official status beyond being a PhD student.

Data Analysis

In analysing the data, I continuously reflected on how my interpretations were shaped by my background and strived to remain objective by using a systematic coding process and discussing the coding tree and framework with supervisors as a form of peer debriefing.

Summary of Reflexivity Statement

Reflecting on my role and biases throughout this study has provided deeper insights into my role as a qualitative researcher. This not only strengthens the rigour of my current research but also informs my approach to future studies, ensuring continued critical self-awareness and ethical mindfulness.

Definitions

For the purposes of this PhD, an **evidence-based guideline** (clinical, public health or social care) follows the definition set out by the US Institute of medicine in 2011:

"Clinical Practice Guidelines are statements that include recommendations intended to optimise patient care. They are informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options. To be trustworthy, guidelines should:

- be based on a systematic review of the existing evidence,
- be developed by a knowledgeable, multidisciplinary panel of experts and representatives from key affected groups,
- consider important patient sub-groups and patient preferences, as appropriate,
- be based on an explicit and transparent process that minimises distortions, biases, and conflicts of interest,
- provide a clear explanation of the logical relationships between alternative care options and health outcomes, and provide ratings of both the quality of evidence and the strength of recommendations; and
- be reconsidered and revised as appropriate when important new evidence warrants modifications of recommendations."(10)

HTA are considered to be a specific type of guideline, typically shorter and more focussed, but following a similar development process.

Qualitative evidence synthesis is "A methodologically rigorous review that uses qualitative methods to synthesise and analyse findings from primary qualitative studies to produce new concepts, understandings and/or theories. The synthesis goes beyond the sum of individual studies by generating new interpretations to which all studies contribute."(11)

In the interests of clarity, this document refers to **guideline committees**. A range of terms are used to represent the group of experts and lay-people who interpret evidence to make guideline recommendations. Examples include guideline panels, oversight committees, decision making committees, panels, advisory boards. The term guideline committees should be interpreted (for the purposes of this work) as encompassing all of those.

Technical staff refers to the people who produce the QES and present the results to the guideline committee, whether they be directly employed staff, contracted staff, or university departments.

Most guideline committees involve people who use the services about which the recommendations are being developed, either as service users themselves or as carers of those people. For the purposes of consistency they are referred to as **lay-people** in this thesis.

Background

The rise (and falter?) of evidence-based medicine/healthcare

Evidence-Based Health Care (EBHC), previously called Evidence-Based Medicine (EBM), has revolutionised clinical practice by integrating the best available research evidence with clinical expertise and patient values. As an approach it aims to optimise decision-making in patient care, ensuring that interventions are effective (12), often including consideration of cost-effectiveness (13). A pivotal component of EBHC is the development and implementation of evidence-based guidelines, which provide health care practitioners with systematically developed recommendations to assist them in delivering high-quality care (14).

One of the consequences of the primacy of EBHC has been a surge in agencies around the world who produce guidelines for health and (to a much lesser extent) social care providers using the methods of EBHC.

Historical Context of Evidence-Based Medicine

In its broadest sense, the roots of EBHC can be traced back to ancient times, where early practitioners based their medical decision-making based on previous observation and experience. However, the most widely accepted dawn of the modern evidence-based medicine era is the work of Professor Archie Cochrane in the 20th century.

In 1972, British epidemiologist Archie Cochrane published ‘Effectiveness and Efficiency: Random Reflections on Health Services’(15), highlighting the lack of reliable evidence supporting many medical interventions. He advocated for the use of randomised controlled trials (RCTs) to assess the efficacy of treatments, laying the groundwork for the development of systematic reviews and meta-analyses that form the mainstay of the modern EBHC movement. Cochrane’s work emphasised the necessity of critically appraising evidence to inform clinical decisions. The actual term ‘Evidence-Based Medicine’ dates back to the early 1990s when it was first coined by Gordon Guyatt and his colleagues at McMaster University(16). They introduced EBM as a new paradigm for medical practice, emphasising the integration of clinical expertise with the best available clinical evidence from systematic research.

Key concepts of evidence-based medicine

Integration of best research evidence with clinical expertise and patient values

EBHC is defined as "the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients"(12). This means integrating individual clinical expertise with the best available external clinical evidence from systematic research. This definition implies that decisions need to be based on the totality of what is known about a topic, an implication that is supported by the existence of electronic databases that promise the possibility of identifying all published evidence on a topic, which can then be sorted and filtered and analysed through systematic processes intended to minimise bias, maximise transparency and provide an answer regarding 'what works'. The role of clinical expertise and patient values in EBHC has often suffered, and to some extent been minimised since Sackett and colleagues wrote about it in the 1990s, to the point where Trisha Greenhalgh and colleagues raised a red flag in the BMJ and called EBM "a movement in crisis" and called for an approach to 'real evidence-based medicine' that focusses on ethical care of the patient, shared decision-making and asking "what is the best course of action for this patient, in these circumstances, at this point in their illness or condition?"(17). While patient and public involvement (PPI) is increasingly promoted within evidence-based healthcare, it is often critiqued for being tokenistic—implemented more to satisfy institutional or funding requirements than to genuinely empower patients or influence decision-making (18). Approaches such as co-production, when properly implemented, offers a more radical and participatory alternative to usual approaches to Patient and Public Involvement (PPI) (19).

Hierarchy of evidence/taxonomies of evidence

EBHC employs a hierarchy to rank evidence-based on methodological rigour and reliability. At the apex are systematic reviews (SRs) and meta-analyses of RCTs, followed by individual RCTs, cohort studies, case-control studies, and case series or expert opinions at the base. Although attitudes are changing, this historical hierarchy of evidence remains at the core of many guideline producers' approaches, with SRs and RCTs being seen as the gold standard methodologies, regardless of the question being asked, or of the methodological quality of the SRs and RCTs themselves. These hierarchies of evidence often do not contain reference

to qualitative research, or of qualitative evidence synthesis, and if they do it tends to sit alongside expert consensus and case series towards the bottom of the hierarchy (see Figure 12 for example).

The positioning of RCTs and SRs of RCTs at the pinnacle of hierarchies of evidence has led to an over-prioritisation of quantitative, experimental evidence, and even though there is an increasing realisation that SRs of RCTs are not always the 'best' evidence, and that RCTs are not always the most appropriate methodology, they continue to be the mainstay of systematic review teams around the world. Researchers have been questioning this reliance on RCTs as the principal means of answering all clinical or healthcare questions since the early 2000s, noting that some questions are best answered (or can only be answered) by examining a range of data sources (20). See the next section for further discussion of this.

Mark Petticrew and Helen Roberts argue that this hierarchy of evidence is often inappropriate for complex interventions (21). They suggest that a rigid hierarchy may not suit all research questions and that a taxonomy of evidence or a typology of evidence might be more useful, where the appropriateness of the method is matched to the type of question being asked. Similar positions are taken by the Social Care Institute for Excellence (SCIE) which promotes evidence-informed practice and supports a broader view of what counts as evidence, including experiential and practice-based knowledge (22), or the Joanna Briggs Institute (JBI) FAME framework—which stands for Feasibility, Appropriateness, Meaningfulness, and Effectiveness (23). This model emphasises:

- **Feasibility**: Can it be done in the real world?
- **Appropriateness**: Is it suitable for the context?
- **Meaningfulness**: Does it resonate with the people involved?
- **Effectiveness**: Does it achieve the intended outcomes?

This approach shifts the focus from just "what works" (effectiveness) to "what matters" in practice. Evidence to decision frameworks such as those used by GRADE, also consider not just the quality of evidence but also values, preferences, feasibility, and equity.

Qualitative research has always struggled to be seen as a legitimate form of evidence for EBHC, and hierarchies of evidence do nothing to assist this lack of acceptance (24). As an

example, a key journal in the medical field, the British Medical Journal (BMJ) announced in 2016 that they would no longer publish qualitative research (25), although they have recently published a guide to thematic analysis (26).

Critical appraisal

A cornerstone of EBHC is the critical appraisal of the selected evidence to assess its validity, impact, and applicability. Reviewers are encouraged to evaluate the methodology of studies, consider potential biases, and determine the relevance of findings to their patient population. Critical appraisal (or quality assessment) attempts to ensure that each trial has been conducted with sufficient rigour. It evaluates the reporting of the methodological detail of the study – how participants were randomised, blinded, recruited etc (27).

Meta-analysis

The preferred approach to synthesising data from RCTs is through pooling the effect estimates of each of the included studies to produce a weighted mean estimate of the overall effect of the intervention on a particular outcome. Such statistical approaches to quantitative data analysis can be highly complex but often have the benefit of generating a single figure 'size of effect' answer. Unlike meta-analyses, which aim to produce a single, quantifiable estimate of effect, qualitative evidence syntheses are inherently interpretive. They involve re-examining and integrating the interpretations of primary study authors to generate deeper, context-sensitive insights that can inform practice and policy. This process emphasizes theoretical development and critical reflection, often resulting in findings that are more nuanced and less definitive (28)..

Evidence-based Guidelines in the UK and their Impact

Evidence-based guidelines are systematically developed statements designed to assist practitioner and patient decisions about appropriate healthcare for specific clinical circumstances (29). They use the results of evidence syntheses, such as meta-analyses, seeking to turn this 'best available evidence' (as detailed above) into actionable recommendations, aiming to standardise care, reduce variability, and improve patient outcomes.

Although numerous guideline-producing bodies, including medical royal colleges and professional bodies, publish guidelines in the UK, the National Institute for Health and Care Excellence (NICE) is the body which decides which treatments are clinically effective and cost-effective enough to be funded by the NHS in England. NICE has established rigorous processes for guideline development (30). This involves:

1. Identifying and prioritising topics based on factors like disease burden, variation in practice, priority for the NHS and potential for improvement in health outcomes.

2. Forming guideline development groups representing multidisciplinary teams, including clinicians, methodologists, and patient representatives to ensure diverse perspectives.

3. Conducting systematic reviews of the best available evidence-based on comprehensive literature searches to gather relevant studies, which are then appraised for quality and relevance, and analysed appropriately.

4. Formulating recommendations based on the evidence, which consider the balance of benefits and harms, quality of evidence, and applicability.

5. Consulting with stakeholders and peer review to ensure accuracy, clarity, and feasibility.

6. Publication.

The implementation of evidence-based guidelines can lead to significant improvements in healthcare (31), including:

- Standardisation of Care: Guidelines provide a framework for consistent practice, reducing unwarranted variations in treatment approaches (32).
- Improved Patient Outcomes: By basing recommendations on robust evidence, guidelines enhance the effectiveness of interventions and patient safety (32).
- Shared Decision Making: Guidelines serve as valuable tools for clinicians and patients, facilitating shared decision making and informed consent (32).
- NHS Resource Optimisation: By recommending interventions with proven efficacy, guidelines contribute to the efficient use of healthcare resources (32).

Challenges in Implementing Evidence-based Guidelines

Despite their benefits, several challenges hinder the effective implementation of evidence-based guidelines. It is a challenge to keep a large portfolio of guidelines up to date in an environment where a vast number of medical and healthcare publications per year are being added to databases. Ensuring that guidelines reflect the latest research findings requires continuous monitoring and timely updates. More recently, as focus has moved towards personalised medicine, it has become more of a challenge to produce guidelines that balance evidence with individual patient needs. While guidelines provide general recommendations, clinicians must adapt them to individual patient contexts, which may not always align with guideline suggestions (33).

The limitations of systematic reviews of randomised controlled trials as the mainstay of evidence-based guidelines

As previously discussed, systematic reviews of randomised controlled trials (RCTs) have long been considered the gold standard of evidence-based medicine, sitting prominently atop the hierarchy of medical evidence (34). However, their limitations and constraints warrant careful examination, as an over-reliance on this methodology may inadvertently restrict our understanding of medical interventions and their real-world effectiveness.

One of the most significant limitations of systematic reviews stems from publication bias in the underlying RCTs themselves. Research has consistently shown that studies with positive results are more likely to be published than those with negative or inconclusive findings (35). A recent study investigating the frequency of publication bias in meta-analyses published in four major general medical journals (BMJ, JAMA, Lancet, and PLOS Medicine) indicates strong evidence of publication bias in 36% of meta-analyses of clinical trials (36). This systematic suppression of negative results creates a distorted evidence base from which reviews must draw their conclusions. Even with comprehensive search strategies and attempts to include unpublished data, systematic reviews cannot fully overcome this fundamental bias in the available literature (37).

RCTs, by their very nature, are conducted under highly controlled conditions with carefully selected patient populations. While this controlled environment enhances internal validity, it often comes at the cost of external validity (38). Systematic reviews aggregating such trials may provide reliable evidence about efficacy under ideal conditions but offer limited insight into effectiveness in routine clinical practice. The strict inclusion and exclusion criteria of RCTs often exclude important patient sub-groups, such as elderly individuals with multiple comorbidities or those taking multiple medicines (39).

Healthcare practices and technologies evolve rapidly, potentially outpacing the timeframe required to conduct and publish both primary RCTs and subsequent systematic reviews (40). By the time a systematic review is published, some of its included studies may already be outdated, particularly in fast-moving fields like oncology or infectious disease management. Furthermore, systematic reviews often struggle to account for temporal changes in standard care practices and evolving treatment protocols (41). This temporal challenge directly impacts the assumption of transitivity in meta-analyses, which assumes that studies are similar enough to be meaningfully compared and that interventions could have been randomised within the same trial (42). When trials are conducted across different decades or contexts, this assumption becomes tenuous – for instance, what constituted 'standard care' in a control group from 2000 may be substantially different from standard care in 2020, undermining the validity of indirect comparisons.

While systematic reviews attempt to synthesise evidence across multiple studies, significant heterogeneity in methodological approaches, outcome measures, and reporting standards can complicate meaningful synthesis (43). Different trials may define and measure outcomes differently, use varying follow-up periods, or employ different control conditions. These methodological variations can make it challenging to draw definitive conclusions, even when statistical techniques like meta-analysis are employed (44).

Modern healthcare increasingly involves complex interventions with multiple interacting components that present distinct challenges for systematic review methodology. Complex interventions are characterised by numerous interacting components, implementation variations, and strong context dependency (45). For example, a hospital-wide infection prevention programme might simultaneously incorporate staff training, new monitoring

protocols, modified cleaning procedures, and updated documentation requirements - all of which interact in ways that can be difficult to disaggregate and evaluate systematically.

Quantitative systematic reviews may struggle to capture and analyse the nuanced effects of these interventions effectively for several interconnected reasons. First, the methodology often oversimplifies the fundamental question from "how and why does this intervention work in different contexts?" to simply "does it work?" (46). This reductionist approach can miss crucial implementation factors that determine success. Second, the substantial heterogeneity in how complex interventions are implemented across different settings creates significant challenges for evidence synthesis. What appears to be the 'same' intervention may manifest quite differently across various healthcare contexts, making straightforward pooling of results potentially misleading (47).

The rigid structure of RCTs and subsequent systematic reviews may not adequately account for the adaptable nature of complex interventions or the importance of implementation context (48). This limitation becomes particularly apparent when considering the tension between intervention fidelity and necessary local adaptation. Healthcare settings often need to modify interventions to fit their specific circumstances, yet systematic reviews often struggle to account for these adaptations whilst assessing effectiveness (49).

Furthermore, conventional systematic review methodology typically focuses primarily on outcomes whilst potentially overlooking crucial mechanisms of action. This approach may fail to elucidate which components are essential, how different elements interact, and what contextual factors are necessary for success (50). For instance, when evaluating a falls prevention programme, a quantitative systematic review might focus solely on fall reduction rates without capturing the critical interactions between staff training, environmental modifications, and assessment protocols that drive success.

The Medical Research Council's framework for complex interventions suggests that more sophisticated approaches are needed, including methods to better describe interventions, understand their mechanisms of action, and account for implementation variations (51). This might involve incorporating qualitative and mixed methods approaches, realist review techniques, and process evaluations alongside effectiveness studies. Recent methodological

developments have begun to address these challenges, though significant limitations remain in current systematic review methodology for complex interventions (52).

Even when systematic reviews include multiple studies, they may still suffer from insufficient statistical power, particularly when examining subgroup effects or rare outcomes (53). This limitation becomes particularly problematic when reviews attempt to draw conclusions about specific patient populations or uncommon adverse events.

In summary, while systematic reviews of RCTs remain a crucial tool in evidence-based healthcare, their limitations necessitate a nuanced approach to evidence synthesis. A comprehensive evaluation framework that incorporates diverse evidence types, including observational studies, qualitative research, and real-world evidence, may provide a more complete understanding of medical interventions and their effectiveness in practice (17).

Moving forward, the EBHC community must work to develop more sophisticated approaches to evidence synthesis that can address these limitations while maintaining the methodological rigour that makes systematic reviews valuable. This might include developing new methodologies for rapid evidence synthesis, incorporating diverse data sources, and better integrating qualitative and quantitative evidence to inform healthcare guidelines and decision making. The role of QES as part of that more sophisticated approach is increasingly being recognised as a key element of a new evidence hierarchy.

The evolution and impact of qualitative evidence synthesis in health and social care research – where are we now?

The growth of QES methods represents a significant methodological advancement in health and social care research, marking a shift from the historical dominance of quantitative systematic reviews. This evolution reflects a growing recognition that understanding complex health and social care interventions requires engagement with experiential and contextual evidence alongside effectiveness data (54). The development of QES methodologies has been particularly crucial in addressing questions about implementation, acceptability, and the lived experiences of both service users and providers.

The emergence of QES has presented a fundamental challenge to strictly positivist approaches to evidence synthesis, which have historically dominated healthcare research.

Positivist paradigms, with their emphasis on objectivity, measurement, and generalisation, have struggled to accommodate the contextual richness and interpretative insights that qualitative research offers (55). QES has helped to legitimise alternative epistemological positions within evidence synthesis, acknowledging that different kinds of knowledge are needed to understand complex social phenomena in healthcare settings (56).

Particularly significant has been the alignment of QES with critical realist perspectives in healthcare research. Critical realism, which acknowledges both the existence of an objective reality and the socially constructed nature of our knowledge about that reality, finds natural methodological allies in qualitative synthesis approaches (57). QES methods support critical realist investigations by helping to illuminate the mechanisms, contexts, and outcomes that characterise complex healthcare interventions. This alignment has proven especially valuable in understanding how interventions work, for whom, and under what circumstances - questions that purely positivist approaches often struggle to address (58).

The epistemological flexibility of QES has enabled researchers to move beyond simple questions of effectiveness to engage with nuanced investigations of causality and context. This has been particularly valuable in understanding complex social interventions, where quantitative systematic review methods may fail to capture important aspects of how and why interventions succeed or fail in different (59). By embracing multiple forms of knowledge and understanding, QES has helped to bridge the gap between positivist and interpretivist approaches to evidence, supporting comprehensive and nuanced understanding of healthcare phenomena.

The emergence of QES can be traced to the late 1980s, with meta-ethnography pioneered by Noblit and Hare providing one of the first formal methodological frameworks (60). However, it was not until the early 2000s that QES began to gain significant traction in health research, driven by growing recognition of the limitations of quantitative synthesis alone in addressing complex healthcare questions (61). This methodological evolution occurred against a backdrop of increasing acceptance of qualitative research in healthcare generally, and growing awareness of the need to synthesise qualitative evidence systematically.

The development of various QES approaches reflects different epistemological positions and analytical needs. Meta-ethnography, thematic synthesis, framework synthesis, amongst others, emerged as distinct but complementary approaches, each offering different possibilities for knowledge synthesis (62). The choice between these methods often depends on the review question, the nature of available evidence, and the intended use of the synthesis findings (63).

Methodological Rigour and Development

Contemporary QES has developed sophisticated approaches to support methodological rigour. The introduction of the ENTREQ guidelines (Enhancing Transparency in Reporting the Synthesis of Qualitative Research) in 2012 marked a significant step towards standardising reporting practices (64). Similarly, the development of quality assessment tools specifically for qualitative synthesis, such as that provided by CASP (Critical Appraisal Skills Programme), helped establish consistent quality standards (65). Both ENTREQ and the CASP qualitative checklist were developed using inadequate tool development methods.. The recent publication of the CAMELOT tool for assessing methodological limitations of qualitative research has in large part superseded previous assessment tools (66), and work is underway in an MRC funded project led by Emma France and Jane Noyes to develop Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) reporting criteria for QES (67).

The methodological sophistication of QES has grown considerably, with approaches now capable of handling various types of qualitative evidence. These developments include:

Theoretical and methodological advancement:

- Best practice in QES is moving beyond simple aggregation of findings to interpretative approaches that can generate new theoretical insights. Meta-ethnography, in particular, has demonstrated its capacity for theory development through the synthesis of multiple qualitative studies (68).
- Mixed methods reviews for integrating qualitative and quantitative data have become increasingly sophisticated, developing methods to integrate qualitative and quantitative findings meaningfully (69, 70).

- Additionally, the development of GRADE-CERQual (Confidence in the Evidence from Reviews of Qualitative Research) has provided a structured approach to assessing confidence in review findings, comparable to GRADE for quantitative reviews (71).

Potential impact on health and social care practice

- The potential influence of QES on health and social care practice is substantial and multifaceted. For example, in:

Understanding implementation: QES has provided crucial insights into implementation challenges and facilitators, helping to bridge the gap between evidence and practice. Studies have shown how qualitative synthesis can illuminate barriers to implementation that might not be apparent from quantitative data alone (72).

Patient experience and engagement: Syntheses of qualitative research have been instrumental in bringing patient voices into evidence-based practice, providing systematic ways to understand and incorporate patient experiences into healthcare decision making (73).

Methodological Challenges and Future Directions

Despite its evolution, QES continues to face several methodological challenges:

Quality assessment: The ongoing debate about how best to assess the quality of primary qualitative studies and their subsequent synthesis remains active, with various approaches proposed but no clear consensus (see above for comments about the CAMELOT tool however) (74).

Integration with quantitative evidence: While mixed-methods reviews are increasingly common, methodological questions persist about the best ways to integrate qualitative and quantitative findings meaningfully (75).

Technology and scale: The growing volume of qualitative research presents challenges for comprehensive synthesis. Emerging methods incorporating machine learning and artificial intelligence may offer new possibilities for handling large volumes of qualitative data (see also below) (76).

A critical examination of the role of QES in developing health and social care guidelines

Contribution to Guideline Development

QES can make several critical contributions to the guideline development process through multiple interconnected mechanisms. The first key mechanism is its ability to substantially enhance the relevance and acceptability of guidelines by incorporating nuanced patient and provider perspectives into their development (77). Furthermore, QES can help to systematically identify and analyse implementation barriers and facilitators by synthesising diverse stakeholder experiences across different healthcare contexts (78).

Methodological Considerations

Approaches to qualitative evidence synthesis in guideline development

The methodological approach chosen for QES is likely to influence its utility in guideline development. For example, the choice between interpretive and aggregative approaches might have substantial implications for how evidence informs guidelines and shapes their eventual implementation (63). Interpretive approaches to QES, such as meta-ethnography and critical interpretive synthesis, focus on developing new theoretical understandings through the synthesis of qualitative findings, for example, about access to healthcare (79). The key strength of interpretive approaches lies in their ability to generate sophisticated theoretical frameworks that explain how and why interventions work in different contexts. However, these approaches typically require more time and resources compared to aggregative methods, potentially limiting their feasibility in some guideline development contexts (80).

Aggregative approaches, for example, meta-aggregation, focus on summarising and combining findings across studies. They can be particularly effective for addressing specific, focussed questions about stakeholder experiences and implementation factors. The primary advantage of aggregative approaches lies in their systematic nature and closer alignment with quantitative evidence synthesis methods. Guideline development teams who are primarily quantitative systematic reviewers are likely to find it easier to adopt aggregative approaches due to their familiarity and perceived transparency. However, these approaches

are less likely to generate new theoretical insights that could inform guideline adaptation across different contexts.

Implications for guideline development

The choice between interpretive and aggregative approaches depends on guideline development needs, resources available and staff skills and training.

Interpretive approaches aim to generate new conceptual understandings by reinterpreting primary data across studies, making them well-suited for exploring complexity and meaning. They are most valuable when:

- Guidelines address complex healthcare interventions requiring theoretical understanding
- Implementation contexts vary significantly
- Available evidence shows substantial contextual variation
- Resource and time constraints allow for in-depth analysis (81, 82)

Aggregative approaches, such as thematic synthesis and meta-summary, aim to summarise and categorise findings across studies without generating new theory, making them ideal for informing policy and practice in a structured and timely manner. This means they are more suitable when:

- Guidelines focus on specific, well-defined questions
- Rapid synthesis is required
- Available evidence is relatively homogeneous
- Resource constraints necessitate a more streamlined approach (81, 82).

A comprehensive piece of work was undertaken as part of the INTEGRATE-EU project (63).

Challenges and Limitations

Despite its value, the integration of QES in guideline development faces substantial operational and methodological challenges.

Time constraints may pose a particular challenge, with comprehensive QES typically taking some time to complete. This timeline often conflicts with policy-driven guideline development schedules, although developing methods for time-sensitive qualitative evidence syntheses are being increasingly used (83). Additionally, guideline organisations often lack staff with specialised qualitative research expertise, leading to potential quality compromises.

Future developments

The field continues to evolve, with several promising directions but also with challenges:

Stakeholder Engagement: There is growing emphasis on involving stakeholders throughout the synthesis process, recognising the value of different forms of knowledge and expertise (84).

Methodological Innovation: Work continues to enhance the methodological rigour of QES, from recent methodological developments such as CAMELOT (previously mentioned) for methodological assessment of qualitative studies, to evidence-based tools to assess data richness and thickness (85). The publication of a Cochrane Campbell handbook for qualitative evidence synthesis (currently largely available online) is also due in February 2026 (86).

New approaches are also being developed to handle complex interventions and context-dependent findings effectively. Realist synthesis approaches, for example, represent important methodological innovations in this space, using both qualitative and quantitative data (87).

Artificial Intelligence, Large Language Models and Automation

Within the field of health and social care generally, but perhaps especially in the area of evidence synthesis, there is unprecedented interest in the benefits and efficiencies that may come from the integration of artificial intelligence (AI), large language models (LLMs), and automation into QES. These technologies present both significant opportunities and challenges for health and social care research synthesis. Over the next five years, they are poised to transform the landscape of evidence synthesis, potentially enhancing efficiency, accuracy, and comprehensiveness.

AI and LLMs can automate labour-intensive tasks such as literature searches, screening of studies, and data extraction. This automation can significantly reduce the time required to conduct systematic reviews, allowing researchers to focus on higher-level analysis and interpretation. For instance, AI-driven tools can quickly scan and categorise thousands of articles, identifying relevant studies in a fraction of the time it would take a human researcher (88). It is unclear whether technologies currently in use at NICE for priority screening (a large language algorithm that re-orders searches based on analysis of human reviewer decisions about inclusion and exclusion during initial sifting) are effective when screening qualitative studies.

AI tools also have the potential to minimise human error in data extraction and synthesis. By standardising processes, these technologies may be useful in checking that data is consistently and accurately extracted, at least in the case of quantitative studies. Whether this will be possible for qualitative studies is unclear.

AI can identify patterns and relationships within qualitative data that may not be immediately apparent to human researchers. These advanced analytical capabilities can uncover new insights and enhance the depth of qualitative analyses. For example, natural language processing (NLP) techniques can analyse textual data to detect emerging themes and trends (88). The key challenge is ensuring the quality and reliability of AI-generated outputs. AI models can inadvertently introduce biases based on the data they are trained on, which can affect the validity of the evidence synthesis. This means that AI tools need to be continuously monitored and evaluated to mitigate this risk and that thorough assessment of the sources of training data and the potential for algorithmic bias is made before these tools are introduced (90).

The use of AI in QES also raises ethical issues related to transparency and accountability. AI processes need to be transparent, and the decision-making criteria used by AI tools must be clearly documented and understood. Ethical considerations also include the potential for AI to perpetuate existing biases and the need for equitable access to AI technologies (89).

Implementing AI and LLMs in QES requires technical expertise and infrastructure that may not be readily available. Training reviewers to effectively use these tools and integrating them into existing workflows can be challenging. Institutions may need to invest in

specialised training programs and technical support to facilitate the adoption of AI technologies (90).

The use of AI and automation involves handling large volumes of sensitive data. Ensuring data privacy and security is paramount. Although systematic reviews and QES do not routinely contain sensitive data related to patients, there are concerns around, for example, storage of copyrighted data outside of the organisation (i.e. on servers belonging to another organisation) (89).

Most importantly, while AI can assist in managing and analysing qualitative data, interpreting the nuanced and context-specific nature of qualitative findings remains a challenge. Human expertise is essential to ensure that the interpretations are meaningful and accurately reflect the data. AI tools are aids rather than replacements for human judgment in qualitative research.

Living Systematic Reviews and Living Guidelines

The following section is a summary of a short paper highlighting the need to consider the future of QES in guidelines if guidelines were to move towards a living model of constant updating. Since the publication of the paper, interest in living guidelines has dampened considerably, however the challenges and opportunities remain. The full proof version of the paper is in appendix L and the published version of the paper is available at:

Carmona, C., Carroll, C. & Baxter, S. The move towards living systematic reviews and living guidelines in healthcare: consideration of the possibilities and challenges for living qualitative evidence syntheses. Syst Rev **12**, 47 (2023). https://doi.org/10.1186/s13643-023-02218-0

The article addresses the challenges and possibilities for developing living QES in response to the growing trend of living systematic reviews and living guidelines in healthcare.

The paper identifies that while QES has become increasingly valued in guideline development for understanding patient preferences and treatment acceptability, the emergence of living systematic reviews and living guidelines —particularly accelerated by the COVID-19 pandemic—potentially threatens the (continued) inclusion of QES. The

fundamental problem is that no established methodologies currently exist for creating "living QES" that can be rapidly and frequently updated.

The commentary outlines key characteristics of living reviews, which include continuous monitoring of evidence, immediate incorporation of new important findings, up-to-date communication regarding review status, and pre-defined decision frameworks for evidence integration. Whilst these protocols are relatively straightforward for quantitative systematic reviews, QES presents unique challenges.

The paper highlights several specific difficulties in adapting QES to a living format. Firstly, searching for qualitative studies is typically more complex than for randomised controlled trials, potentially requiring greater resources for regular screening. Secondly, different QES methodologies may encounter varying levels of difficulty in incorporating new evidence. Interpretive approaches (e.g., meta-ethnography) might require substantial reworking to accommodate new data, whereas aggregative approaches may integrate new findings more readily.

A distinctive consideration for living QES is the currency of qualitative data. Unlike quantitative evidence on drug efficacy, which may remain relatively constant, qualitative evidence reflects social and individual views that evolve over time. This raises questions about when evidence should be "retired" from a living QES and whether such syntheses have finite lifespans before requiring complete revision. The paper also address how living QES might inform living guidelines. Unlike quantitative reviews, where changes in effect size may trigger guideline updates, the implications of new qualitative evidence may be less clearly defined. The commentary concludes that QES methodologists must urgently develop efficient processes for updating QES if the synthesis of qualitative evidence is to continue meeting the needs of health guideline producers in this emerging era of living evidence.

This piece represents a call to action for qualitative researchers to ensure their methodologies remain relevant and valuable in the rapidly evolving landscape of evidence-based healthcare, where the demand for current, frequently updated guidelines is increasing.

Timeliness and the Move to Short Updates of Recommendations

Internationally, guideline producers and systematic reviewers are facing pressure to produce guidance quickly. For example, the WHO approach to rapid guideline development during emergencies has evolved since the Ebola outbreak in 2014. Since then, they have developed living guidelines for COVID-19 clinical management and therapeutics in response to urgent public health needs. The WHO Guidelines Review Committee implemented specific methodology for "rapid advice guidelines" that significantly compressed previous timelines (91). The Scottish Intercollegiate Guidelines Network (SIGN) has implemented expedited review processes for urgent clinical questions. During the COVID-19 pandemic, SIGN partnered with NICE on rapid guidance development, with their joint rapid guideline on managing COVID-19's long-term effects demonstrating this accelerated approach (92), and Cochrane established specific methodology for rapid reviews, allowing evidence synthesis in 1-2 months instead of 1-2 years (91). The Cochrane COVID-19 Study Register was created to streamline identification of relevant studies, supporting their rapid review initiative during the pandemic (93).

Currently NICE is implementing a wide-sweeping transformation. Part of the current transformation plan is a focus on the timeliness of guidance products (https://www.nice.org.uk/about/who-we-are/corporate-publications/the-nice-strategy-2021-to-2026), which in real terms means that they have the rather naively paradoxical aim of producing guidance more quickly while maintaining rigour. Currently, internal teams at NICE of 1.3 to 2.3 full-time equivalent staff produce a quantitative systematic review in 6-8 weeks. In 2014, the average time to produce a systematic review (based on registry data from PROSPERO) was 67.3 weeks for a team of average five people (94), although developments in automation tools have led to more recent phenomena like the 2-week systematic review process (95), along with wider acceptance of ‘rapid review’ methods such as those recently updated by the Cochrane Rapid Reviews Methods Group (96). Producing high-quality QES can be a slow process, involving substantially more reflection and critical engagement than a typical systematic review process, and recent developments in methods for QES are remarkable in their comprehensiveness and adherence to the tenets of the qualitative paradigm. However, as a result they are quite slow procedurally. Examples could include the Interactive Summary of Qualitative Findings (ISoQ) tool (97) for GRADE-CERQual,

which is the interactive software for assessing the level of confidence in themes resulting from QES, but requires extensive setting up, or the recently released CAMELOT tool (66) for assessing the methodological limitations of qualitative evidence for use in QES. While far superior to the previously favoured CASP checklist (98), the CAMELOT tool may be more time-consuming to complete.

In spite of this, there is a developing momentum to understand what a 'rapid' QES might look like, and the Cochrane Rapid Reviews Methods Group and the Cochrane Qualitative and Implementation Methods Group published a paper in early 2024 describing some of the possibilities. They identify methods of synthesis that work well with large numbers of studies or amounts of data and conclude that "judicious use of Grading of Recommendations Assessment, Development and Evaluation approach for assessing the Confidence of Evidence from Reviews of Qualitative research assessments and of software as appropriate help to achieve a timely and useful review product." (99).

A somewhat related point where a rapid QES process would be perceived as beneficial is in the changes to the way that NICE manages its portfolio of guidelines, with an emphasis on shorter updates to specific recommendations or to specific small areas of a guideline rather than full updates of older guidelines. In many cases, the trigger for these updates is the publication of a new trial or effectiveness study that has been monitored by the NICE surveillance team. However ongoing qualitative studies are not monitored and therefore an update would never be triggered by a qualitative study that had been monitored being published.

Summary

The growth of QES represents a significant methodological advancement in health and social care research and has had a large impact on EBHC. Its evolution from relative obscurity to mainstream acceptance reflects both methodological maturation and growing recognition of the importance of qualitative evidence in understanding complex health and social care interventions, however, its place is far from secure. As the field continues to develop, maintaining methodological rigour while embracing innovation will be crucial for its continued evolution and utility. One of the areas where development is needed concerns

exactly how it might be used in guideline development. Chapter 2 explores this from a historical perspective by analysing existing guidance on using QES in guideline development.

CHAPTER 2: SYSTEMATIC REVIEW OF THE METHODOLOGICAL LITERATURE FOR INTEGRATING QUALITATIVE EVIDENCE SYNTHESES INTO GUIDELINE DEVELOPMENT

This chapter presents the first study carried out as part of the PhD. It comprises a systematic review of the available methodological literature to explore and summarise what is known about best practice in incorporating evidence from QES in the development of health guidelines.

The primary aim of this chapter is to identify gaps in the reviewed literature where there is little analysis or guidance on parts of the process of using QES to inform guidelines, and to identify those areas of the process where there is broad consensus about methods, and those areas where there is either no consensus or disagreement.

This chapter reports the first study undertaken for this PhD, which means that some time has passed since this review was completed. Most of the important work published since then is reported in the previous chapter, and, prima facie, none would have met the inclusion criteria for this review. However, a full re-run search was not undertaken.

An abridged version of this chapter was published in Research Synthesis Methods in 2021 (see appendix E for final proof version).

> Carmona C, Baxter S, Carroll C. Systematic review of the methodological literature for integrating qualitative evidence syntheses into health guideline development. *Res Syn Meth*. 2021; 12: 491–505. https://doi.org/10.1002/jrsm.1483

Within this chapter general references are listed in the reference section with Vancouver style referencing. Studies formally included in the analysis are referenced Harvard style to allow consistency of attribution while reading. The full references for these papers are listed at the end of the chapter.

Introduction

The use of findings from qualitative evidence syntheses as evidence in the development of health guidelines is growing as the need for relevant and context-sensitive evidence increases (100). This is commonly agreed to be because qualitative data can answer

particular types of questions far better than quantitative data. Quantitative data is still key for studies of efficacy, but the parts of clinical effectiveness that are dealt with by understanding patient preference, and other softer contextual outcomes such as feasibility and acceptability can best be brought to light by qualitative study (101). The move towards QES is also driven by the move towards greater patient-centredness in health systems, for example shared decision-making, and the greater inclusion of patients and lay experts in guideline-producing committees (102).

Lewin et al. (103) build on the theme by claiming that incorporating QES into guideline development can help to represent people who may otherwise be excluded from the process (p.3) and Carroll (104) notes that QES can also potentially offer a valuable supplement to the experiences of patient representatives on guideline panels.

NICE (105) give some examples of when QES might usefully answer questions related to guideline production (p.78):

- How do different groups of practitioners, people using services or stakeholders perceive the issue (for example, does this vary according to profession, age, gender, or family origin)?
- What social and cultural beliefs, attitudes or practices might affect this issue?
- How do different groups perceive the intervention or available options? What are their preferences?
- What approaches are used in practice? How effective are they in the views of different groups of practitioners, people using services or stakeholders?
- What is a desired, appropriate, or acceptable outcome for people using services? What outcomes are important to them?
- What do practitioner, service user or stakeholder groups perceive to be the factors that may help or hinder change in this area?
- What do people affected by the guideline think about current or proposed practice?
- Why do people make the choices they do or behave in the way that they do?

- How is a public health issue represented in the media and popular culture?

This does not mean that advocates of QES in guideline development are oblivious to the challenges of the approach. Several consistent concerns arise about using QES within guideline-producing bodies:

Training – Authors acknowledge that guideline producers are principally systematic reviewers who may have no background or expertise in qualitative methods, and therefore need training to be able to produce high-quality QES (106).

Methodological issues –Many qualitative researchers do not support the notion of synthesising qualitative research, and for those that do there is no universally accepted way of doing this in health and social care (102) (though this position has changed somewhat since the publication of that paper). Qualitative research itself is criticised as being context-dependent and specific, for including an insufficient number of informants, for being interpretative and, because it usually relies on small, purposive samples, for having a low degree of generalisation (107, 108). An additional issue here is the linking, mixing, or merging of qualitative and quantitative evidence. As Carroll notes, there is no ready-made toolkit for doing this (104). Additionally, when Tan et al. (109) reviewed the use of qualitative data by NICE up to 2009, they noticed a lack of consistency in terminology and method and even lack of agreement about what constituted a qualitative study across their different guideline-producing centres (p.172)

Committee processes – Lewin et al. (103) and Glenton et al. (101) refer to processes about how committees make decisions on the basis of qualitative evidence, and furthermore, how the strength of evidence relates to the strength of recommendation when QES is included, with Glenton et al. noting that the WHO guidelines have had their guidelines criticised for making 'strong recommendations' despite there being only low or very low confidence in the underpinning evidence. They point out that WHO guideline panels are expected to take into account broader evidence about acceptability, feasibility, and equity, in addition to evidence about effectiveness.

Given the increasing pressure to use findings from qualitative research in the development of health guidelines, alongside the ongoing concerns over their use and the readiness of

guideline-producing bodies to integrate this kind of evidence into their processes, it is crucial to examine the methods that are being adopted or proposed both by experts in the field, and by guideline-producing bodies themselves. Recent growth and development in methods and standards for qualitative evidence synthesis, and the development of tools to ensure transparency both in qualitative synthesis (for example CERQual (110)) and in guideline development (for example, the DECIDE collaboration Evidence-to-Decision framework (111) has put qualitative evidence firmly on the agenda for evidence-based medicine, but the most appropriate methods for using it during guideline development remain unclear. This chapter will systematically review the methodological literature that addresses this topic with the aim of producing a thorough overview of the state of the field.

Review question

What methods and processes are proposed in the methodological literature for incorporating QES into evidence-based health guideline development?

Protocol

The protocol for this review is in appendix A

The protocol was written using the fields laid out by PROSPERO, the international database of prospectively registered systematic reviews where there is a health related outcome (112). This review protocol was not prospectively registered with PROSPERO because it does not currently accept registrations for scoping reviews, literature reviews (such as this one) or mapping reviews. However, the PROSPERO framework for review protocols provides a useful and comprehensive framework to ensure that the key relevant parameters of the review have been planned.

Methods

Inclusion and Exclusion Criteria

Health related databases were searched for papers published in English since 2000 that met three criteria:

- They were about the synthesis of primary qualitative evidence
- They were about a process of guideline production

- They described some level of methodology for using the QES to inform the development of guidelines

These criteria were chosen for several reasons. Firstly, searching for qualitative health studies is challenging in itself (113), and searching for qualitative methodological papers even more so. Methodological reviews typically involve sensitive searches across multiple databases (114) which, as a result of the sensitivity, retrieve large numbers of papers. This is resource intensive and involves high volume of sifting for little yield. Secondly, recent years have seen an explosion of publications describing general QES methods (73, 115) and a further generic review of QES methods would have added little to this literature.

On this basis it was agreed that limiting the searches to papers that reported methods of QES specifically in the context of health guidelines would be a way to manage the large volume of papers that would need to be sifted.

Search Strategy

Searches were conducted across the following health databases:

- MEDLINE (including MEDLINE in-process)
- EMBASE
- CINAHL
- PsycINFO

From 2000 up until 15 August 2019

Supplementary searches were conducted in Google Scholar (116) using the search string ‘qualitative synthesis guideline development’. Relevant results from the first eight pages were added to the search results.

Reference lists of included papers were checked for further potential includes.

A full search history can be found in appendix B.

Results from searches were downloaded into EPPI reviewer software (117), where they were de-duplicated in preparation for sifting.

Sifting

Titles and abstracts for all papers were screened using the EPPI reviewer 5 software. Papers that appeared to meet the three inclusion criteria outlined above, or those where it was uncertain whether the criteria were met or not were identified for further examination. The papers identified as potential includes were downloaded in full text form and a second round of sifting took place using the full text of the articles. Articles marked for inclusion at this stage also had their reference lists checked to identify further papers.

To maximise transparency of selection at the full text stage, a simple checklist was used to check that papers met the criteria for inclusion as described above (see appendix C for completed checklist).

Data Extraction

Included papers were uploaded in portable document format (pdf) into NVivo 12 (118) and were marked up using a combination of a priori and emergent codes that related specifically to stages in the reviewing process.

A simple a priori data extraction structure was set up to map to different stages of the reviewing process. The purpose of the structure was to enable the sorting of the content of included papers into the different stages of the review process to allow comparison and aggregation of the content of the papers by review stage. The stages were chosen to represent the discrete tasks that are involved in creating a classical systematic review or QES. They were specifically kept at a general level because early reading of the included papers indicated that there was insufficient granularity in them to justify a more nuanced structure:

- Protocol/scope/review question
- Search
- Study selection
- Data synthesis
- Critical appraisal

- Quality appraisal
- Making recommendations
- Use of logic models/frameworks
- Integration with quantitative data/reviews
- Reporting

Along with three additional categories:

- Benefits of using QES in guidelines
- Challenges of using QES in guidelines
- Existing QES methodologies that have been used in guideline development

Additional emergent themes were coded as 'child' codes as they occurred to allow for additional stages of the guideline development process to highlight themselves during the data extraction process.

Results

Flow of Papers

Systematic searching of databases (appendix B) yielded a total of 5,822 references. These were uploaded into EPPI reviewer 5 software (117) and a deduplication algorithm was run. 756 duplicate records were identified and removed from the database.

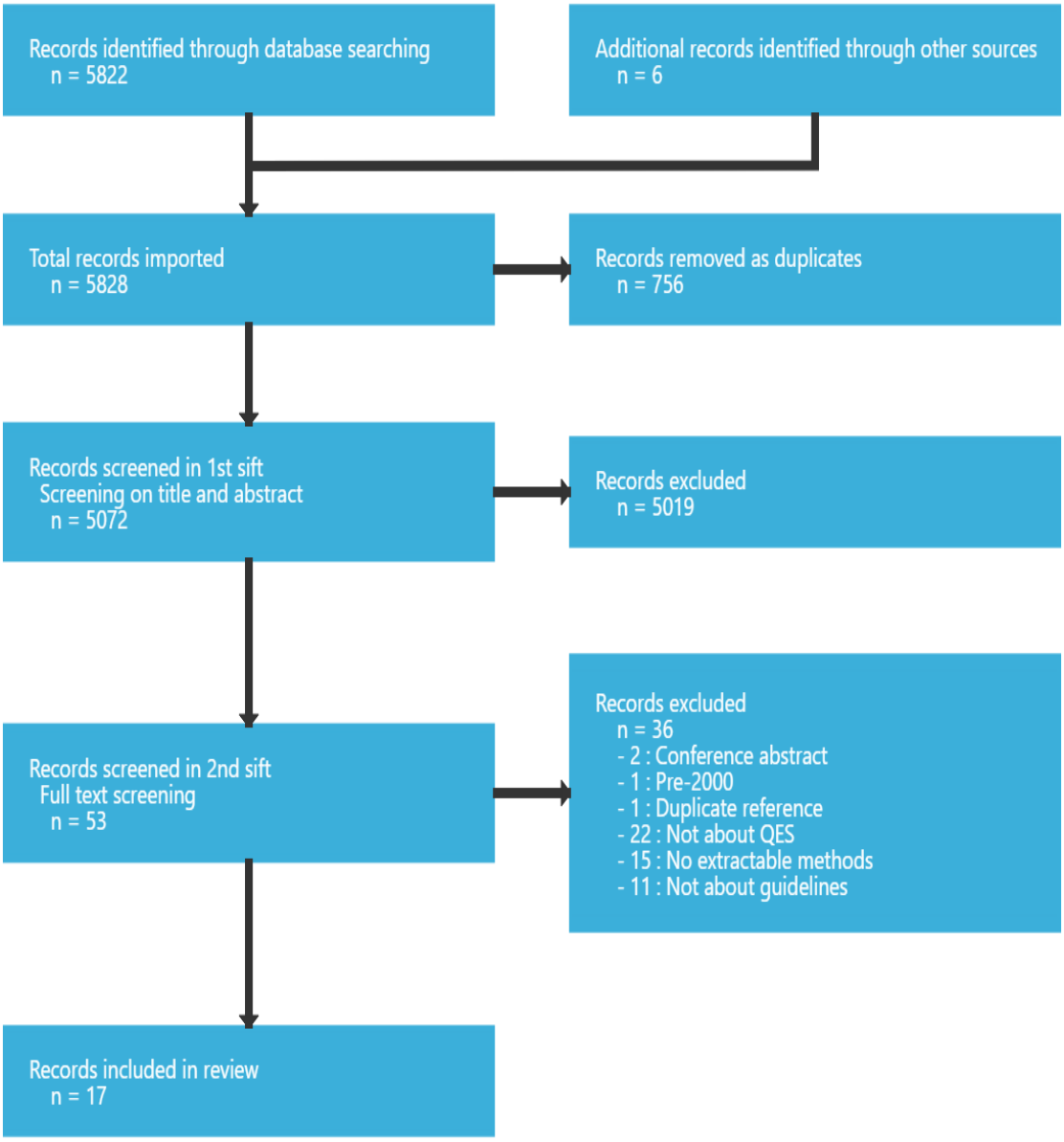
In total, 5,066 records were screened at title and abstract level. Records that stated or implied that the full paper was about qualitative evidence synthesis, guideline development and had some methodological detail about a process for using QES in guideline development were marked for full text assessment. 5,019 records were excluded with 47 papers marked for full text examination.

Forty-seven papers were assessed against a study inclusion checklist (appendix C) to check whether they met the above inclusion criteria. Papers that did meet the inclusion criteria had their reference lists checked for additional relevant articles. Eleven articles met the

inclusion criteria, and six additional references were identified from reference list checking. A total of 17 papers were included in the review.

For a PRISMA diagram describing the flow of papers, see Figure 1.

Figure 1: PRISMA flow diagram



Excluded papers

For a list of papers excluded at full text, along with reasons for their exclusion, see appendix D.

Included Papers

Full references of included papers are provided at the end of the chapter.

Table 1 provides a brief summary of each paper.

Table 1: Brief details of included papers

Author	Type of paper	Summary
Booth 2016	Guidance	A report funded by the EU as part of a series on evaluating complex interventions ('integrate-HTA'). The guidance document sets out a framework to enable reviewers to choose between different QES methods depending on the question they are asking.
Carroll 2017	BMJ Analysis	The analysis focusses on the need for successful guidelines to reflect patient views and argues that qualitative evidence is a key way to do this. The paper is not primarily a detailed methodological paper but contains some extractable methodological detail.
Downe 2019	Research article	The first in a series of three papers that have been written by a group of methodologists working with the WHO on guidelines that integrated QES. The authors examine the use of QES in developing clinical and health systems guidelines.
Flemming 2019	Analysis	This paper presents an overview of the ways QES can be used to address complex interventions.

Author	Type of paper	Summary
Glenton 2016	Manual/handbook	Chapter 15 of the WHO handbook for guideline development specifically about using evidence from qualitative research to develop WHO guidelines.
Glenton 2019	Research article	The third in a series of three papers describing the use of QES to inform the development of clinical and health systems guidelines by a team of methodologists who have worked with WHO. The WHO is increasingly using evidence derived from QES to provide information on acceptability and feasibility in its guidelines.
Gould 2010	Methodological report	Gould describes qualitative work done to support the production of two social care guidelines by the UK National Institute for Health and Care Excellence (NICE).
Hansen 2011	Methodological report	This article focuses on qualitative research synthesis in eliciting patients' perspectives as part of the growing drive to include patient views in policy and HTA.
Knaapen 2015	Toolkit (chapter)	Chapter 2 of a Guidelines International Network (GIN) toolkit on patient and public involvement in guidelines. It contains practical ideas about how to conduct a qualitative evidence synthesis as part of the guideline development process.
Kristensen 2007	Manual/handbook	The 2007 updated edition of the *Health Technology Assessment Handbook* that was issued by the Danish National Board of Health in

Author	Type of paper	Summary
		2001 as part of the fulfilment of the National Strategy for HTA. Contains some general detail about QES and also a specific chapter on assessment and syntheses [sic] of qualitative studies (s.4.2).
Lewin 2018	Commentary	Argues that the development of more robust (transparent) methods and tools for QES has widened the opportunities for QES to be used to inform health guidelines (in the context of the WHO).
Lewin 2019	Research article	This is the second in a series of three papers written by methodologists working with the WHO that examines the use of QES in developing clinical and health system guidelines. It specifically discusses using qualitative findings as part of Evidence-to-Decision frameworks.
NICE Manual 2018	Manual	The process manual used by NICE to produce clinical guidelines. The NICE manual includes details of synthesis for all the types of evidence it uses, not just qualitative evidence.
Ring 2010	Guidance	Guidance from NHS Quality Improvement Scotland about the various methods of QES that could be used in HTA.

Author	Type of paper	Summary
Ring 2011	Research article	The authors conducted a systematic search to identify QES and reflect on the methodological approach used.
Swedish Agency for Health Technology Assessment and Assessment of Social Services (SBU) 2016	Manual	Swedish agency for health technology assessment manual for evaluating and synthesising qualitative material.
Tan 2009	Evidence utilisation report	Describes the use of qualitative research as evidence in a national clinical guideline program (National Institute for Health and Clinical Excellence – NICE, UK).

Final Data Extraction Framework

The 17 papers were uploaded in.pdf format to NVivo 12 as described above. During data extraction several additional 'child' code nodes were created where it was possible to add granularity to the a priori nodes. The final data extraction structure is presented in Table 2. The nodes to which the data was extracted are in column 1, column 2 indicates the number of papers that contributed data to that node, and column 3 gives the total number of items of data from those papers extracted to that node, for example, node 1 'Review protocol and scoping' has 11 items of qualitative data for seven different papers. Node names that are inset and bulleted are 'child nodes' of the node above.

Table 2: Data extraction structure from NVivo - methods and processes proposed in the methodological literature for incorporating QES into evidence-based health guideline development

Name of node	Papers that contribute to node	Total items coded to node
1 Review protocol and scoping	7	11
• Reflexivity	2	2
2 Identifying evidence	12	29
3 Study selection	3	7
4 Data synthesis	9	22
5 Critical appraisal	10	13
6 Reporting standards*	3	3
• Evidence statements	1	1
• Evidence tables	2	2
• Reporting tools	1	1
• SoQF[1]	1	1
7 Recommendations*	3	9
• CERQual	6	10
• Committee	3	5
• EtD frameworks[2]	5	22
○ Implementation issues	1	7
• Using frameworks	1	3
Benefits of using QES	7	10

Challenges	9	9
Methodologies for QES	8	21
Overall process	3	5
When to use qualitative reviews	1	1

* In this analysis child codes were not aggregated into the parent node, so the parent node does not contain all of the data in the child nodes.

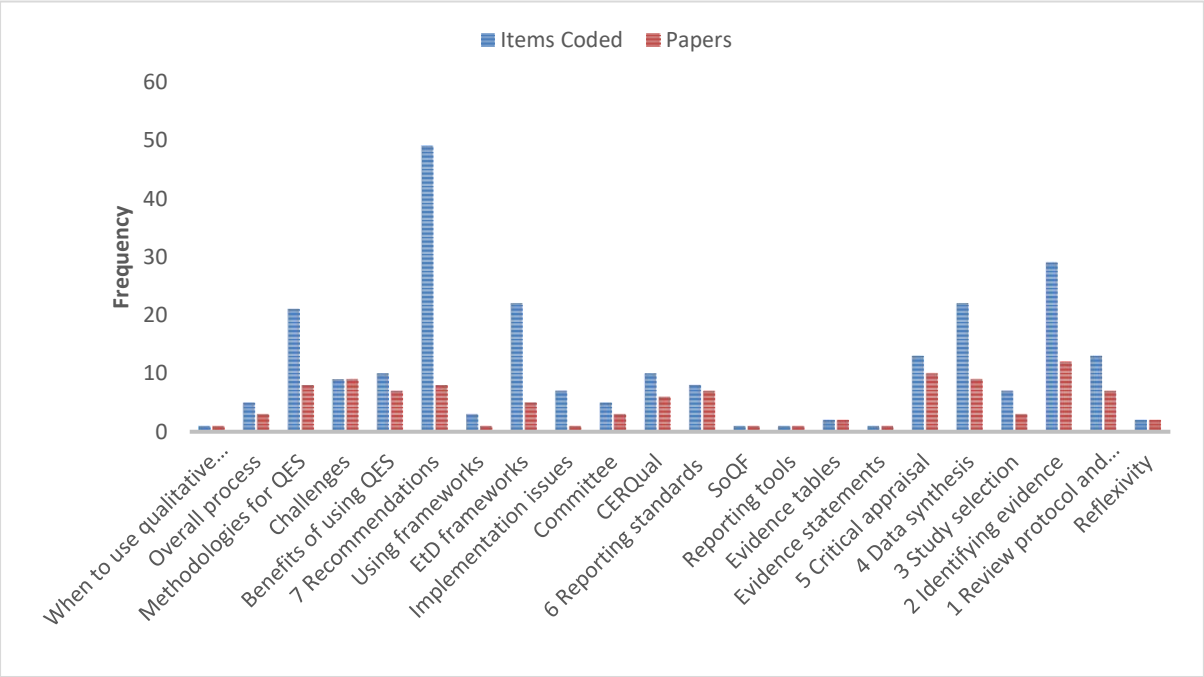
[1] SoQF – summary of qualitative findings

[2] EtD framework – Evidence-to-decision framework

Distribution of Data

The distribution of data extracted across the 17 papers was very variable, with the distribution of data clustering more closely around the areas of QES where there is greater consensus, and with fewer authors tackling the issues that are perceived as more difficult. A breakdown of the distribution of coded data is presented in Figure 2.

Figure 2: Coding frequency



Description of extracted data

The following sections aggregate the data extracted from the 17 included papers by node (as described above). Where the data are consistent enough to do so seamlessly, nodes have been reported under the same subheading. The methods and processes proposed in the methodological literature for incorporating QES into evidence-based health guideline development are as follows:

Methods for review protocol development or scope generation

Seven of the 17 included papers discussed or referred to the need for a scoping process or a process of review protocol generation before the searching and literature identification phase of the development of a guideline. As part of this preparation phase, two studies (Downe et al. (2019) and Ring et al. (2010)) make reference to the importance of reflexivity. It is not normal practice in quantitative systematic reviewing to provide any kind of reflexivity statement because quantitative research designs are intended to minimise the effects of bias through randomisation and robust and transparent methods. Qualitative primary research, however, tends to embrace reflexivity and positionality, recognising the researcher's influence as an integral part of the inquiry rather than a bias to be removed.. Downe et al. characterise a reflexivity statement as a way to be transparent about the views, values, and beliefs of the review authors so that the readers of the review have an "insight into the lens through which the authors have viewed their data.". Both papers gave an example of a reflexivity statement.

In terms of the scoping or review protocol development itself, Lewin et al. (2019:15) describe the critical tasks for the scoping phase of a guideline as identifying the interventions, stakeholders, and contexts relevant to the guideline questions. They note that this can be time-consuming, and that adequate time needs to be set aside for this part of the process.

In general, the scoping or protocol development phase was regarded as the first phase of the development of the guideline and was a precursor of any qualitative or quantitative synthesis. Knaapen et al. (2015:28) note the importance of considering the review questions before beginning to review the literature, and the scoping or protocol development stage is the first step towards doing this.

However, Glenton et al (2016:184) point out that qualitative evidence can also be a useful way to establish the scope of the work to be undertaken itself, suggesting that a QES might be used to inform the scope of the guideline and that there might be a more dynamic relationship between the guideline committee and the scope that could be at least partly mediated by QES.

The protocol/scope itself should make use of expert input to ensure that it is relevant. This could be topic experts or policy makers to ensure the review question is topic/policy-relevant (Ring et al.2010), or it could enlist the help of the guideline development group to ensure relevance (Glenton et al.2016), though this process is time-consuming and may slow down the protocol development process (Ring 2010). It should include the objective of the review, criteria for including studies the search strategy, data collection and analysis, and a reflexivity statement. The criteria for including studies can be based on a modified PICO(S) framework or can make use of the SPICE protocol (SBU 2016:17), though the authors note that its isn't necessary that every component is part of every study. The framework can be regarded as heuristic, as a "helpful guide".

Ring et al (2010) propose that the review protocol or scope should be made publicly available before the review commences in the same way as a quantitative systematic review would be registered. QES protocols can be registered in international databases of prospectively registered systematic reviews such as PROSPERO (112).

Overall, there is broad agreement that a review protocol or scope should use expert input (for example a guideline committee, or service user organisation) to derive the review question and the criteria for that review (for example using a PICO or SPICE format). The protocol should be the first stage of the process unless there is some search for evidence to inform the scope or protocol (for example a QES). Scopes or protocols for QES may include a reflexivity statement.

Methods for identifying literature

Two thirds of the included papers contained information regarding the optimal methods of identifying evidence for a qualitative evidence synthesis. Most agreed that as part of the process there needed to be a systematic search of databases and pointed out that in many

ways this was similar to quantitative database searching. Like literature searching for quantitative reviews, there is a focus on systematic and reproducible searching across a range of databases that suitably cover the area of the research question. This may involve going beyond the usual medical and healthcare databases, especially if the question relates to public health or social care (Hansen et al. 2011:146 and Gould 2010:103), and it is suggested that relevant additional databases to be searched include CINAHL, PsycINFO, Applied Social Sciences Index and Abstracts (ASSIA), Social Science Citation Index (SSCI), and AnthroSource (Knaapen et al. 2015). The choice of databases should be driven by the research question (SBU 2016:19). This focus on systematic searching represents a step-change from earlier in the 2000s where searching seems to have been less developed. A review of qualitative evidence used in the production of NICE clinical guidelines by Tan et al. (2009) found that at that point in time there was no evidence of any standardised methodology across the 22 clinical guidelines that incorporated qualitative evidence. Most appeared not to have done specific searches for qualitative literature (or didn't report it) in specific qualitative/social sciences databases, and only four of the 22 guidelines reported specific methodological detail for the qualitative studies they included.

In spite of their overall support for systematic searching, several authors were quick to point out that searching for qualitative studies is more complex than for quantitative studies. Ring et al. (2010) summarise the main reasons for this as inadequate indexing in databases (although this is improving), non-indicative wording in titles and abstracts (for example the use of a participant quote as the main title of the paper), and that the focus of the qualitative study may not reflect the research question of the qualitative synthesis but the paper may still contain relevant data.

While NICE (2018) and Ring et al. (2011) both argue for systematic searching, authors note that it may not be important to identify every available study, citing theoretical saturation as a possible endpoint. Ring et al. (2011:388) additionally note that QES conducted alongside a systematic review will be more likely to have explicit inclusion criteria than a synthesis of qualitative studies that aims for theoretical saturation where there might be a more iterative approach to searching and screening.

It is considered important by Downe et al. (2019) to do preparatory work to expedite the searching and to ensure that the search process is robust. Ideas for doing this include undertaking scoping searches to test the strategy. This can help to identify potential concepts and values that may help frame the guideline.

SBU (2016) and Lewin (2019) recommend searching for any existing QES. Even if none of the QES found meet the protocol for the QES at hand, they may still be useful for informing the search strategy. Lewin et al. add that an increasing rate of QES being published makes it increasingly likely that relevant QES may be found that address some or all of the guideline. Kristensen (2007:49) recommends the development of a specific search protocol to inform the searches and for the purposes of transparency.

Purposive sampling is frequently used as an adjunct to, or occasionally as a replacement for, exhaustive systematic database searching. Purposive sampling is described by Knaapen et al. (2015) as an iterative process of searching and screening, with the process being complete when the reviewers achieve “theoretic saturation” [sic] or “conceptual robustness”.

The NICE methods manual (2018) states that “it may not be necessary to identify all the literature on a topic. The objective may be to reach theoretical saturation, where any additional studies identified merely support the existing line of argument, rather than identify all relevant studies. In this context, it may be possible to undertake searches which are more precise.” (p. 90). It does not give any detail on how those more precise searches might be undertaken.

The use of filters to restrict the volume of data returned from searches is recommended by SBU (2016) and Knaapen et al.(2015:30), however, in spite of the existence of validated filters, searching for qualitative literature can lead to extremely large searches with a low specificity, which in turn leads to large volumes of irrelevant literature to be sifted and rejected (Ring et al.2010:12). They argue there is a balance to be struck between the time-consuming sensitive searches and more specific ones that are likely to miss some studies.

As a result of concerns over missing data for the reasons above, most authors recommend some kind of additional search method. Methods for additional searches that were mentioned include:

- Footnote and reference list checking for included studies
- Detailed reporting of searching in publications to allow future researchers to duplicate and develop the search methods used
- Hand-searching key journals relating to the topic of the QES
- Forward citation searching (searching for relevant work by locating studies that cite earlier key studies)
- Author searching (searching for all publications by the author of a relevant work)
- Contacting authors and other key informants to provide references
- Searching the grey literature, although papers recommend searching grey literature, none of them imply any kind of systematic search of the grey literature but rather propose a 'berry picking' approach to find the easily identifiable literature.

Downe et al. (2019) also refer to three other considerations relevant to identifying data. They recommend considering the date range (if a specific range of dates is relevant to the review); the types of publications that will be included (for example, what methodologies and methods of data collection are included? Are mixed methods studies relevant? Will grey literature, commentaries or unpublished studies be included?) and the language of publication, though since the paper is written in the context of QES used to develop WHO guidelines, it could be argued that the emphasis on multiple language use is less relevant to QES that are aimed at producing national or regional guidelines.

In summary, while most authors advocate some kind of systematic searching process, the complexities of identifying qualitative literature have led to approaches that try to reduce the volume of literature found by comprehensive searches (for example by using filters or more specific search terms), but then trying to mitigate the potential loss of relevant papers by adding in supplementary search techniques such as citation searching.

Methods of study selection

Only three of the 17 papers gave any detail about considerations required in selecting qualitative studies. Downe et al. (2019) and Ring et al. (2010) at some length, and SBU

(2016) in a single paragraph. Fundamentally the authors agreed that the process for study selection of qualitative literature should mirror the process that would be expected in a quantitative systematic review, with multiple reviewers comparing the paper with pre-specified inclusion and exclusion criteria, with any disagreements resolved by discussion or by the use of a third reviewer. This was especially the case when QES was being conducted alongside a quantitative systematic review (Ring et al, 2010:9).

Ring et al. (2010) and Downe et al. (2019) highlight that study selections should be transparently reported, for example through using a PRISMA diagram (Preferred Reporting Items for Systematic Reviews and Meta-analysis (119)) to show the flow of studies through various stages of selection.

Downe at al. (2019) warn about the possibility of retrieving a large number of studies and recommend that in these cases reviewers select a sample. They note that "a 'large number' is difficult to quantify and will, to a certain extent, depend on the emerging themes and concepts as well as the resources available and the timeframe required to complete the review. Reviewers should seek to ensure that no one sampling system affects the overall quality of the review by introducing reviewer bias." (p.6).

Methods of quality appraisal

Whether or not there should be some form of critical appraisal, or quality appraisal of qualitative studies included in QES is an ongoing discussion within the broader field of QES, and those using QES for guideline development are no strangers to those same arguments. Arguably, researchers undertaking QES for the purposes of guideline development are more often in favour of transparent and systematic methods than researchers with a pluralist approach to QES. It may be for this reason that overall, the ten included studies that discuss critical appraisal are broadly in favour of using some method of critical appraisal to assess the methodological quality of the studies included in a QES. This position is likely to be strengthened by the introduction and widening use of the GRADE-CERQual tool (110) for assessing confidence in findings from QES since CERQual relies on a methodological assessment (amongst other things) of the studies included in the review finding (Knaapen 2015:32).

Tan et al. (2009:171) report there is a lack of consistency in the use of qualitative checklists in qualitative reviews for NICE clinical guidelines. Only half of the guidelines with qualitative research reported any form of critical appraisal, and of the 11 that were critically appraised, there was variation in the method used to do that, using checklists from CASP (the Critical Appraisal Skills Programme), SIGN (the Scottish Intercollegiate Guidelines Network) and SCIE (the Social Care Institute for Excellence).

The NICE manual (2018) states unequivocally that "Critical appraisal of qualitative evidence should be based on the criteria from the Critical Appraisal Skills Programme" (p. 106) and this is echoed by both Glenton et al. (2016) in the WHO handbook and the SBU methods manual (SBU, 2016). Downe et al. (2019:6) are also unequivocal that "the included studies should be subjected to a formal quality appraisal using one of the recognised appraisal systems for qualitative research". Hansen et al. (2011) share a similar view in relation to HTA. The authors report that "[their] experiences of HTA processes used on patients' perspectives and [their] reading of the debates in the literature lead [them] to support the view voiced by some researchers that the quality of the studies must be assessed systematically to avoid drawing unreliable conclusions." (p.146).

Other authors are more cautious and refer to lack of agreement about the value of critical appraisal of qualitative studies. Ring et al. comment both in 2010 in the NHS Quality Improvement Scotland report and in their 2011 journal article on this. They argue that some checklists are "reductionist and over-prescriptive in nature" (2010:12) and that there is currently little consensus over what constitutes a 'high-quality' qualitative study (2011). In spite of this, they note that the EPPI-Centre, CRD and the Joanna Briggs Institute (JBI) all assess the quality of qualitative studies. The 2011 report (p.389) also links analytic strategy to quality appraisal, and notes that thematic synthesis, for example, has a particular approach to critical appraisal whereas some other approaches are less wedded methodologically to critical appraisal.

Knaapen et al. (2015:33) succinctly summarise the same issue – "The use of standardised 'checklist' approaches has been strongly critiqued by some commentators, questioning how quality criteria modelled on the principles of positivist science can be applied to non-positivist qualitative research" (p.33). However, in spite of this they provide a summary of

the strengths and limitations of a range of checklists, including the CASP tool (98), Chapter 21 of the Cochrane manual (120), Spencer et al.(121) and QARI (122).

It is also noteworthy that some specific approaches to QES (for example framework synthesis and thematic synthesis) have their own approaches to critical appraisal, whereas in other approaches such as meta-ethnography or grounded theory critical appraisal is less important (Ring et al. 2010:13).

Methods of synthesis – General approaches

Much of the description of methods for the synthesis of findings from primary qualitative studies was presented as methods for specific QES methodologies. These specific approaches will be discussed after an outline of the generic methods referred to by other authors.

The Swedish HTA handbook for evaluation of qualitative studies (SBU 2016:24) describes the evidence synthesis process as having four discrete stages. Firstly, papers are read to give an overview of themes, then the papers are read and coded. No detail is given on the method of coding, but the manner in which it is described implies a process of emergent coding where codes are allowed to emerge from the included papers. As a second stage these 'first level themes' themes are "distilled to form the second level theme". This appears to be an aggregative coding process of drawing together similar codes. Thirdly, an interpretive coding phase, described by the paper as follows - "Related second level themes are finally synthesised to an overall third level theme. Important patterns and associations among the second level themes are interpreted and problematised. The process is repeated until third level themes are set". The final stage in their synthesis process is "a general assessment of the scientific basis is made. Thereafter evidence graded results and conclusions are formulated", which appears to describe some assessment of review findings, like CERQual or GRADE.

This process of descriptive and interpretive coding is also referred to by Flemming et al. (2019) who note that the need for descriptive or interpretive themes is dependent on the nature and purpose of the QES, which is in turn dependent on the nature of the guideline being developed. The outputs from framework syntheses or thematic syntheses are often as

simple as a list of themes identified across included studies with little or no interpretation that can be used to "detail the needs, values, perceptions, behaviours and experiences of stakeholders within the guideline" (p.7). On a related note, Lewin et al. (2019:4) suggest that QES used in guidelines tends to focus either on people's views about the interventions under scrutiny by the guideline, or it relates more widely to people's views and experiences of the condition underlying/addressed by the intervention(s) that the guideline is examining.

Later in the paper Flemming et al. (2019) add that if interpretive findings are being produced then there is a need for transparency on the part of the reviewers to ensure the interpretations are plausible and to show how they were arrived at (p.8).

Knaapen et al. (2015) also agree that findings should be 'added up' or compared and contrasted, though the process of doing so inevitably masks their variability (p.34). This makes it easy to lose sight of the individuality of participants and their context that are the very heart of qualitative research. They exhort reviewers to strive for this and advise that "[s]uch a process combines the 'distilling down' of individual studies (into summaries and evidence tables) to reduce diversity, with the creation of 'remainders' where the differences, details and contexts of the original studies is preserved (in appendices and footnotes.)" (p.34). Downe et al. (2019:9) follow a similar line of argument about the need to balance between splitting themes emerging from synthesis to the point that they are no longer useful and lumping data together into themes that oversimplify or lose variation in the data.

Downe et al. (2019) make specific reference to the use of Evidence-to-Decision frameworks (EtD) as a driver for the style of the QES. For example, they report

> *The main purpose of an EtD-orientated QES is to generate a series of findings from the included data, which are directly focussed on interventions addressed in the guideline, assessed for confidence and tailored towards acceptability, feasibility and equity, and the values that stakeholders attribute to the outcomes associated with the intervention. The findings are then added to the guideline EtD frameworks, prior to guideline panel consideration, (p.8)*

and Lewin et al. (2019:5) pick up a similar thread in a different paper in the same series, noting that the use of infographics and logic models incorporated into EtD frameworks in cases where the synthesis is intended to be explanatory or theory building.

Only Ring et al. (2010:12) and Knaapen et al. (2015:34) mention the use of CAQDAS (Computer-assisted (or aided) qualitative data analysis software) such as NVivo (118)as a means of organising the data. Knaapen et al. (2015:34) also mention ATLAS.ti (123), MAXQDA (124) and QARI, which is part of the Joanna Briggs Institutes System for the Unified Management, Assessment and Review of Information (SUMARI) software (125).

Overall, discussion about the general methods to be used for synthesis focuses on the continua between aggregative or integrative coding and interpretive coding and between lumping and splitting of themes. This depends to some extent on the methodology used for the analysis as described in the next section.

Methods of synthesis - Specific methodologies

Several authors provide brief (or occasionally in-depth) descriptions of methods of synthesis that can be used. It is not the remit of this chapter to reproduce general methodological detail about the various methods, but where authors have made comment on what makes a method suitable or unsuitable to produce a QES for a guideline development process, that has been included here.

There are a range of different QES methodologies available, some more developed than others. They predominantly reflect methods of primary qualitative research. The different methodologies sit broadly on a continuum between aggregative (or integrative) approaches that summarise themes, and interpretive approaches that generate new interpretations of the data (Flemming et al. 2019:4). They also point out that the Cochrane Qualitative and Implementation Methods Group recommend that the method of synthesis should only be chosen after the pool of evidence for the review is known and caution against pre-specifying a methodology (p.5).

Selection of a method is seen by authors as complex and dependent on many factors, especially the distinction between aggregative methods (where themes are integrated/aggregated) and interpretive methods (where the researchers try to add

additional layers of interpretation over the data). Ring et al. (2011:386) cite the philosophical view of the researcher and purpose of the review as driving factors, whereas Carroll (2017:2) highlights the possibilities of "pragmatic and relatively rapid methods of qualitative evidence synthesis" that might fit better with guideline developers' timelines. He also comments that "Framework, narrative, and thematic synthesis are particularly useful for answering questions about the uptake of interventions and for integrating quantitative and qualitative findings. These methods are therefore potentially the most appropriate for use in developing clinical guidelines" (p.1) and the paper reports that NICE already uses some form of thematic synthesis in some of their public health guidelines.

This identification of thematic synthesis methods is in line with the latest iteration of the NICE methods manual (2018), which continues to identify thematic analysis as an appropriate methodology for analysing qualitative data. It advocates extracting 'first level themes' into evidence tables. (Evidence tables are detailed summaries of the content of each study included in a review or synthesis. These are normally incorporated into an appendix of the review or synthesis.) and using those to generate 'second level themes' in the body of the synthesis. The manual also goes on to discuss (in passing) conceptual mapping, grounded theory, meta-ethnography, and meta-synthesis, but notes that expertise in their use is needed.

Most of the studies that specify methodologies refer predominantly to the same pool of methods – Hansen et al. (2011: pp.147 onwards) give some brief details about several synthesis methods: meta-synthesis, 'imported concepts', meta-ethnography, meta-study, and qualitative meta-summary. Knaapen et al. (2015:34) give very brief coverage of meta-summary, meta-synthesis and meta-ethnography and the framework approach. The older Danish HTA manual (Kristensen et al. 2007:65) mentions only meta-ethnography and narrative synthesis. This is likely because limited QES methodologies were available at that time. Ring et al. (2011:386) note that when they surveyed 107 different reviews that synthesised qualitative data, they found that reviews using critical interpretive synthesis, meta-interpretation, qualitative cross-case analysis, and grounded theory synthesis were found infrequently, and that therefore their usefulness as methods of synthesising qualitative research for HTA is unknown.

The most comprehensive and well-developed guidance for selecting an appropriate QES method for HTAs is the report for the INTEGRATE-HTA project (Booth et al. 2016). This project develops various criteria for QES and matches them to 19 different methodologies. Users of the guidance can compare the various methods for conducting QES to their needs in comprehensive tables that clarify a diverse range of considerations for each method. The project was directed specifically at HTA methods, but there seems no reason to suppose they would not be equally applicable to broader health guidelines.

Overall, thematic synthesis is the most frequently mentioned form of QES in guidelines and seems to be the most commonly used with meta-ethnography and framework or best-fit framework synthesis as alternatives. This is primarily because other methods have not been well tested, so they may be useful or not (Ring et al. 2011). Booth et al. (2016) demonstrate that while a broad range of QES methodologies can be useful, the art is in selecting the appropriate methodology for the research question and research context. This resonates with the assertion by Flemming et al. (2019:4) that the method of synthesis should only be once the pool of evidence for the review is known.

Reporting standards

Reporting standards for QES were not discussed at great length in any of the included papers, possibly because different organisations have well established reporting standards internally. Carroll (2017:2) does note that "…generic qualitative evidence synthesis reporting guidelines exist, others are being developed for particular methods, and standards are evolving to establish the level of confidence users can ascribe to the findings of such syntheses."

Flemming et al. (2019) and Downe et al. (2019) both allude to the importance of systematic review-like transparency in QES. Flemming points out that historically transparency has not been handled well by people reporting QES but highlights work that has been undertaken to develop reporting standards for QES, such as the ENTREQ tool (64) and the eMERGe tool for meta-ethnography (126).

Downe et al. (2019) report what seems to be a useful minimum reporting standard from their work with the WHO. The standard matches closely with the reporting standards for

quantitative systematic reviews (Cochrane reviews particularly) and suggests the characteristics and critical appraisal of each study should be presented in some detail, accompanied by a summary of themes (summary of qualitative findings) along with the confidence in those review findings along with the reasons for any downgrading. They also suggest a list of excluded studies along with reasons for exclusion (p.9).

The guideline handbooks also briefly recommend approaches to reporting, with the WHO handbook (Glenton et al.2016:192) recommending the use of a summary of qualitative findings table that includes CERQual assessment (if there is one). SBU (2016) report quite cryptically that

> *After the quality assessment process, the studies are tabulated and then stratified according to method/research design. In cases where the raw material comprises text which reflects what the informant has said, allocation to category should be exemplified by quotations. Other cases, for example observation studies or action research, require a presentation of how the categories in the synthesis have been formed, or if they are based on the original categories in the included studies. In synthesising the results of studies conducted according to different research methods, caution is required, and the choices made should be discussed and justified. Thereafter the synthesis is initiated, which means that the results from the different studies are combined to form new perspectives or views. (p.24)*

The authors do provide tables and examples that clarify somewhat how their evidence would be laid out.

The NICE manual (NICE 2018) is more prescriptive and requires researchers to provide evidence tables for all included studies that show "bibliography (authors, date) study aim, study design and setting (for example, country) funding details (if known) population or participants theoretical perspective adopted (such as grounded theory) key aims, objectives and research questions; methods (including analytical and data collection technique) key themes/findings (including quotes from participants that illustrate these themes/findings, if appropriate) gaps and limitations overall comments on quality, based on the critical appraisal and what checklist was used to make this assessment." (p.113)

When a summary of qualitative findings is not being used, the NICE manual requires the production of evidence statements – "Evidence statements for qualitative studies or synthesis of qualitative studies do not usually report the impact of an intervention on behaviour or outcomes, and do not report statistical effects or aggregate measures of strength and effect size. Instead, statements should summarise the evidence, its context and quality, and the consistency of key findings and themes across studies (meta-themes). Areas where there is little (or no) coherence should also be summarised." (p.117)

Moving from evidence to recommendations

Papers discussed various aspects of the process of evidence-based recommendation-making that fall generally into four categories:

- Certainty in findings (including CERQual)
- Frameworks (including Evidence-to-Decision frameworks)
- Committees
- Making recommendations from the evidence

Certainty in findings from QES

Lewin et al. (2019) state that "an assessment of confidence in or certainty of the evidence is required by a number of guideline development agencies, including WHO, to ensure that those making recommendations can take into account both the review finding and information on confidence in that finding" (p.2)

To all intents and purposes, at least for papers published after its release in 2015, this means using GRADE- CERQual. Since its publication, CERQual seems to have become somewhat ubiquitous when using qualitative evidence for guideline development. Lewin, Glenton and several of the authors collaborating in the 2019 series of papers for the WHO series (Lewin et al. 2019, Glenton et al. 2019, Downe et al.2019) were part of the original team who authored and devised the CERQual system for assessing the certainty in findings of qualitative evidence, and in the WHO papers they recommend the use of CERQual in guideline development.

The earliest mention of CERQual in the included papers is in the WHO guideline handbook qualitative chapter (Glenton et al.2015:191), which contains a brief description of the components of CERQual as a tool to measure the level of confidence, in each of the findings of the QES. They also note its similarity to GRADE for qualitative studies.

NICE (2018) recommend the use of CERQual somewhat more robustly. The manual notes that unless the qualitative evidence is very sparse or very disparate (in which case a narrative approach is appropriate), the results of QES should be presented as summaries ('at outcome level') and should be assessed with GRADE-CERQual (p.238). They present 'evidence statements' (narrative summaries) as a less preferred alternative (p.113)

Frameworks

The three papers in the recent WHO series (Downe et al. 2019, Glenton et al. 2019 and Lewin et al. 2019) discuss at some length the use of Evidence-to-Decision frameworks (EtD), an approach developed by the GRADE working group to increase transparency in moving from evidence and contextual considerations to implementable recommendations (128, 129). These EtD frameworks take the form of tables that draw together the key information necessary for guideline committees to make recommendations, including the PICO for the research question, summaries of the evidence, details of equality issues, feasibility issues, implementation consideration etc. They contribute to the overall transparency of the movement from evidence through discussion by a guideline committee or similar into recommendations but are not specific to QES. Lewin et al. (2019) refer to the evidence from QES being added to the evidence section of the EtD framework alongside any quantitative evidence, along with its CERQual assessment (p.7), however they also note that "The nature of this type of evidence [QES] means that it does not always fit well within the summary-based and compartmentalised structure of the EtD framework. This may also be an issue where the technical team use findings from QES that were not undertaken specifically for the guideline" (p,5). Glenton et al. (2109) discuss the implementation issues related to clinical guidelines and advise that evidence from QES that does not make it into the evidence section of the EtD framework can often be rewritten a little and turned into an implementation consideration (p.9).

The only author outside of the three WHO papers who discussed EtD frameworks was Flemming (2019:2) who mentions that "[a] QES can be conducted separately or can be integrated with some form of quantitative synthesis. Within a guideline development process, findings from a QES will often be integrated with evidence of effectiveness in an evidence-to-decision (EtD) framework, used to formulate recommendations" (p.2).

In a broader discussion about frameworks generally, Glenton et al. (2019) highlight the usefulness of frameworks in organising data for a QES, and also for identifying gaps in qualitative data.

Committees

There was surprisingly little discussion of the role of any kind of guideline or oversight committee in interpreting the evidence generated by QES and using it to develop recommendations, beyond the discussion reported above in relation to EtD frameworks.

The primary reason for this lack of discussion is probably related to a lack of understanding about the processes by which committees use evidence to generate recommendations – Lewin et al. (2019) point out that "We also do not yet have a good understanding of how guideline panels use and adjudicate different types of evidence (quantitative, qualitative) addressing different types of questions (effectiveness, feasibility, etc.) in making a decision". (p.15)

Gould (2010) discusses a process undertaken in the production of a social care focussed guideline (on dementia) where the evidence was searched for and reviewed by an academic review team, but the "weighting and synthesis" of evidence was done jointly with a guideline committee that included patients and carers (p. 101). Knaapen et al. (2015:35) suggest that committee members may need to be reminded of relevant qualitative evidence, and "while any group member may be expected to read, mobilise, integrate and value its findings, this championing role might more easily be taken up by the producer of the synthesis, the methodologist or patient representatives." (p.35).

Making recommendations from the evidence

The process of making recommendations using the results of QES was not discussed in-depth in any of the papers, with those that mention it mostly reporting that it is difficult to

capture by simple steps and rules (Knaapen et al.2015:34). Normally, committees (in whatever form they take) make recommendations based on the one or more systematic reviews, including any QES, alongside any other information that the committee consider to constitute 'evidence' (for example EtD frameworks). However, Glenton et al. (2019) warn that this is not always the case and points out that sometimes confidence in QES or other types of evidence-based on published studies may be overridden, for example by human rights considerations or other overarching principles or normative values (p.10).

So, although the evidence is primary, it is not the only consideration for guideline committees, and that is true of both quantitative and qualitative syntheses, according to Lewin et al. (2019:14), the amount that any kind of evidence drives a decision about a particular recommendation should depend on the question being considered, and the judgments made should be supported by clear and transparent justifications.

There is little contained in the included papers to explain or clarify the process of using QES to inform recommendations.

Discussion

Overall, the literature relating to the use of QES within the context of guideline development seems to mirror large parts of the general literature on QES, and this is of little surprise since the key people driving the development of QES methods in health and social care are also often the same people who are driving the agenda for using QES in guideline development processes.

In places where there is variation in approaches to QES, the robust approach (from a positivist viewpoint) tends to become the advocated approach where guideline production is concerned. In a world where evidence-based healthcare is dominated by the systematic review of randomised controlled trials (RCTs) by organisations such as NICE, Cochrane, WHO etc, a model of QES that matches their already existing standards of methodological robustness is likely to be more acceptable to them (and to fit better with their exiting methods of interrogating evidence). The Cochrane Qualitative and Implementation Methods Group (QIMG) have been instrumental in this, most notably by supporting the development of GRADE – CERQual for assessing the level of confidence in summary qualitative findings in a way that clearly (and purposefully) matches the process of using the GRADE tool on

quantitative pooled outcomes, but also by supporting methods of QES that can be integrated into or presented alongside Cochrane systematic reviews.

Overall, in this sample, there is a level of excitement about the possibilities of using syntheses of qualitative evidence alongside quantitative evidence. Lewin and Glenton (130) summarise this in their paper that cautiously proclaims a 'new era' for qualitative research, supported by recent developments in QES methodologies such as robust methods for synthesis and for assessing the confidence that can be placed in the findings.

In spite of their enthusiasm, there is still not universal agreement that qualitative evidence can be synthesised in a way that is meaningful or useful. This is mainly because primary qualitative evidence makes no claim to be generalisable, yet for a QES to be useful to a guideline-producing committee, the committee need to be able to argue that the evidence is generalisable enough that it speaks to common experience. Although these arguments seem to be broadly ignored by researchers producing QES currently, the early days of QES were dogged by these arguments (62). As Booth notes in the introduction to his PhD thesis (131), "[QES] has emerged from the confluence of conventional systematic review methods with methods for primary qualitative research. With such a mixed heritage, and the juxtaposition of quite different epistemological positions, it is inevitable that the resultant tensions have generated considerable creative energy and significant methodological frictions."

Once past the epistemological arguments about the philosophical feasibility of QES, the practical methods of performing a QES seem to have converged in terms of their applicability to guidelines at least over the past decade. There seems to be broad agreement over most stages of producing a QES to inform a guideline, even if the fine detail is not always consistent; but there is a lack of clarity on exactly how the resulting QES is used (or not), hence the importance of the work in Chapters 4 – 6, which investigate this. The following section highlights similarities and differences laid out in the results section:

Review protocol: Review protocols are considered important in driving QES, perhaps using SPICE, PerSPEecTiF or SPIDER rather than modifying quantitative PICO formats. It may be beneficial to involve lay-people and experts in protocol development, but this is resource intensive and time-consuming. Different frameworks for formulating research

questions/protocols are developing as a response to growing demands for robust methods in QES, for example recent work on the PerSPEcTiF framework (132).

None of the included studies mentions a role for tertiary reviews in presenting qualitative evidence to committees, presumably because there are not sufficient published QES to make a 'synthesis of syntheses' plausible. The use of tertiary 'reviews of reviews' is established in the quantitative literature (for example, the Cochrane handbook (120) Chapter 22), and is often used for scoping reviews or mapping reviews to provide an overview of the field.

There is also an underlying assumption throughout these papers that a single QES would underpin an entire guideline. Only one paper (Lewin et al 2019) raises the possibility of multiple QES for a single guideline, where different QES are relevant to different questions, and could potentially use different methodologies to synthesise the evidence.

Identifying literature: There is substantial agreement that database searching for literature through structured searching using validated qualitative filters is generally a good idea, however because this often misses studies that are poorly indexed or whose title does not obviously match their content, some kind of supplementary searching is also common. This can be reference list searching, citation searching, asking experts, or trawling grey literature. There is some support for introducing a concept of theoretical saturation into literature searching to prevent it becoming too onerous. Once the searches are not giving us any new themes (in the case of iterative searching) then stopping searching is justifiable. This has a parallel in software for screening quantitative evidence that uses priority screening algorithms to order papers according to what has been previously selected. This allows screeners to stop screening once no more includable papers have been found for a pre-specified time.

Although theoretical saturation may be a robust approach, it remains difficult to operationalise as it relies on an interaction between the searching and the synthesis components (often done by different teams in a guideline-producing process). For a recent example of developing methods in this area see Ames et al. (133).

For organisations producing guidelines who have an interest in capturing the broadest possible population (for example, to be able to make recommendations relating to inclusion health), using theoretical saturation to stop searching or including new data risks missing important data that would be useful in terms of being able to produce recommendations for different sub-populations of the target population.

Virginia Braun and Victoria Clarke (134) have been critical of the concept of theoretical saturation in qualitative research, particularly in the context of reflexive thematic analysis. Their main criticisms are:

Incompatibility with Reflexive Approaches

Theoretical saturation is rooted in positivist assumptions about data completeness and objectivity, which are at odds with the interpretive and constructivist foundations of reflexive thematic analysis. In their view, meaning is not simply discovered in data but co-constructed through the analytic process, making the idea of "saturation" problematic.

Illusion of Objectivity

Saturation implies a finite and discoverable truth within the data, which contradicts the idea that qualitative analysis is inherently subjective and shaped by the researcher's lens. This can lead to a false sense of methodological rigour.

Constraining Creativity

The pursuit of saturation may limit analytical creativity and depth, encouraging researchers to stop data collection prematurely or to treat analysis as a mechanical process rather than a dynamic and iterative one.

Misuse as a Sample Size Justification

Saturation is often used to justify sample sizes in a formulaic way, rather than being grounded in the goals and nature of the specific research project.

A key concern in searching the literature (that also impacts on study selection below) is that relevant data may be included in studies that are not directly relevant to the research

question at hand. Therefore, searches may need to be more sensitive than for quantitative data, meaning that potentially much larger numbers of studies need to be screened.

Study selection: Little was discussed relating to study selection; however this is probably because the study selection process mirrors that in quantitative studies – the studies are matched against the inclusion and exclusion criteria in the review protocol, usually in a two-stage process, firstly based on title and abstract, and then for papers that are not obviously excludable, at full text.

Quality appraisal: General agreement about the importance of using some kind of transparent process for quality appraisal is greater within the QES for guidelines literature reviewed for this study than among the general QES literature. Again, this is probably most likely to be a mirror of the importance of robust process and transparency in systematic reviews of the quantitative literature. There is little agreement about the best way to measure the quality of a qualitative study because of the methodological variation between different styles of analysis and differing views about what a 'good' quality study looks like. The importance of quality appraisal has reached a head since the CERQual assessment became available – one of the components of CERQual is 'methodological limitations' as assessed by a critical appraisal tool. A recent review (135) found 102 critical appraisal tools for qualitative research as the first stage of developing a new tool called CAMELOT and the publication of the CAMELOT tool in 2024 has the potential to resolve this concern. In the meantime, the least provocative tool seems to be the one developed by the Critical Appraisal Skills Project (CASP), and that is the one most frequently suggested by papers in this review. A further challenge to the appraisal of qualitative studies concerns whether or not studies of very low quality should be excluded from any analysis or whether in spite of their poor quality they can contribute to the overall analysis.

Synthesising findings: The different approaches to synthesising data are the most variable area of the review presented here. Different authors either propose their preferred method, or in some cases, attempt to present an overview of the different key methods that are used. Several things remain unclear following the review. It is uncertain whether some methods are more appropriate than others for particular types of question. This seems to be the view held by the publication for the INTEGRATE-HTA project (63), which presents a

range of support to help reviewers choose between different methodologies depending on the research question to be answered. Several authors recommend meta-ethnography (60). For several of the older studies this is likely to be because the method was published in 1988 (the first published method for QES) and was the most established method at that time. More recent papers include a wider variety of methods, but the most common (and probably the most accessible) seems to be some form of thematic analysis that goes through a process of aggregative coding, either followed or not by a stage of interpretive coding. It remains to be shown whether multiple methods are the best way forward, or whether a single method is most useful (both pragmatically in terms of its speed and ease, and methodologically in terms of the usefulness to guideline committees of its output); and of course, if a single method is most appropriate, which one?

Authors who have reviewed the use of different methods are clear that most methods of QES have not been used often in guideline processes, and therefore it is unclear whether they are useful.

Reporting: The lack of detailed discussion of reporting standards in any of the included papers is likely to be in large part due to the fact that most of the authors were writing from within centres where reporting standards are already developed by the organisations that housed them. This is true of NICE (NICE 2018, Tan et al. 2009) and is likely to be true of WHO (Glenton et al. 2016, Downe at al. 2019, Glenton et al. 2019, Lewin et al. 2019), SBU (SBU 2016) and the Danish Centre for Health Technology Assessment (DACEHTA) (Kristensen et al. 2007). The ENTREQ tool (64) is commonly cited, and many published QES adopt some variation of the PRISMA standards for quantitative reviews (119). The Cochrane review format, adapted for qualitative reviews, is also a useful standard for presentation.

Recommendations: It is unclear how committees move from QES findings to making recommendations, and much work done in recent years has been an attempt to make this process more transparent. This is the case for both quantitative studies and for qualitative studies. The advent of GRADE and GRADE-CERQual have been valuable in this, but there is still not always an obvious link between the review findings (and any other information considered by the guideline-producing bodies as evidence) and recommendations made by the committee (and the relative strength of the recommendation). A recent analysis (of

GRADE related recommendations) concluded that this was primarily because panel (committee) members were not adequately trained and did not understand the GRADE system (136). WHO seem committed to using DECIDE-GRADE Evidence-to-Decision frameworks as a way of making the issues the committee need to consider clearer; however, the papers imply that there continue to be challenges in this approach. There does not appear to be an extant evaluation of the usefulness of EtD frameworks to health guideline committees, but the issue of transparency and consistency in recommendation-making is clearly a critical factor in credible guideline production.

Gaps highlighted by the review findings

This review has highlighted some areas where there appears to be enough consistency between different approaches to give a reasonable level of confidence in them. In some cases, there seems to be better agreement between authors writing about QES methods for guideline production than for QES more generally, for example in terms of the level of agreement around searching and around the need for critical appraisal of qualitative studies, both of which are contentious in the broader field of QES (74).

There are also areas where there is less clarity, some of which are likely to be quite specific to QES in the context of guideline production and therefore could probably not be resolved by wider searching of the literature.

As more and more QES are published in health and medicine, it becomes more likely that reviewers will find existing QES that wholly or partially answer their research questions. There is no discussion in the literature to explore how these may be used. Parallels in quantitative systematic reviewing include updating and using pre-existing reviews as evidence for committees; other reviewers use the inclusion lists from systematic reviews as a check that they have identified the relevant literature. In areas where several very similar systematic reviews exist a ‘review of reviews’ or tertiary review can be conducted. None of those things are reported in the papers included here, but similar methods might be possible for QES.

Much of the literature included in this review contains the unspoken assumption that one guideline will require one QES. Only one paper moots the possibility of multiple QES for one

guideline (Lewin et al. 2019), however it is easy to imagine a guideline that contained questions that could be informed by several QES. There is no discussion in the included papers of how this might work in practice.

In terms of producing QES that are useful to guideline committees as part of the evidence base they consider, is a standardised methodology best or is methodological pluralism more useful where the methodology can be a more pragmatic choice and take into account the time, resource and outputs that are wanted? Booth et al. (2016) identify some 30 methods for conducting a QES, and Noyes et al. identify a further ten methods that are in development (137). Few of those methods have been used frequently in producing QES for guideline development and so their utility is uncertain.

A related gap in this review is an understanding of what the most useful way is to provide guideline committees with the outputs of various reviews, both qualitative and quantitative alongside other kinds of evidence. In the past, guideline producers prioritised reviews of RCT evidence to provide evidence of clinical efficacy, but the hierarchy of evidence for other types of outcomes is less established. As QES become more robust in their methods and transparency, can they become the priority evidence for certain types of guideline question? One series of papers discusses in some depth the use of EtD frameworks, but these have not been the subject of robust evaluation, and it is unknown how useful they are to committees

Finally, there is little research, and none in this review that explores how committees move from QES findings and the other information they are given as evidence – systematic reviews, expert testimony, real-world data etc, in the context of their own expertise and experience - to making decisions that produce guideline recommendations, and this is a fundamental question for guideline producers.

Limitations of the Review

As with most reviews, somewhat artificial boundaries have to be placed on the scope of the review to make it manageable and internally coherent. It is rare for a topic, especially for a methodological review, to be clearly circumscribed. This review is no exception.

The content of this review was specifically limited to papers describing methods for using QES in guideline production. As a result, it does not cover the large, and growing, corpus of

literature dealing with the topic of QES generally. There are good published overviews of the development of QES, notably through leadership from the Cochrane Qualitative and Implementation Methods Group (138). Fortunately, several of the key authors writing about the use of QES in guideline development are also key authors in the field of QES in health more generally, so the literatures largely inform and reflect each other.

For the purposes of time and resource, this review did not consider the implications of the upswell in methods for integrating qualitative and quantitative data to produce mixed-methods reviews. As technologies for producing robust QES and robust SR develop, researchers are becoming interested in integrating qualitative and quantitative data in the hope that the whole may be greater than the sum of the parts.

Conclusions

The use of qualitative evidence syntheses to inform the production of health guidelines is growing as the methods for producing them become more clearly defined and more standardised. Methods for producing QES for guideline committees tend to be similar to quantitative systematic review methods in terms of searching, appraisal of evidence, systematic management of data and presentation of results. While this allows greater transparency and greater accountability, it could be argued that it is less 'true' to the principles of being 'led by the data', which are fundamental to most qualitative research. Led by the data means that researchers approach the data without imposing rigid frameworks, allowing patterns and meanings to emerge organically. It reflects a commitment to inductive reasoning, where insights are grounded in participants' experiences rather than predetermined theories (139).

Recent developments in QES mean that there is broad agreement about how QES can be produced to help inform guidelines, but further research is needed to establish whether guideline-producing committees find QES useful to their deliberations, whether they could be done or presented differently to make them more useful and, perhaps most importantly, how committees use QES to inform their decision-making alongside quantitative systematic reviews of effectiveness. This last point is addressed in the following chapters.

Included papers

This list only references studies included in the review. General references are included in the full reference list.

Booth, A., Noyes J, Flemming K, Gerhardus, A., Wahlster, P., Van Der Wilt, G.J., Mozygemba, K., Refolo, P., Sacchini, D., Tummers, M., Rehfuess E (2016) Guidance on choosing qualitative evidence synthesis methods for use in health technology assessments of complex interventions. [Online]. Available from: http://www.integrate-hta.eu/downloads/.

Carroll, Christopher (2017) Qualitative evidence synthesis to improve implementation of clinical guidelines. BMJ (Clinical research ed.) 356: j80.

Downe, Soo, Finlayson, Kenneth W, Lawrie, Theresa A et al. (2019) Qualitative Evidence Synthesis (QES) for Guidelines: Paper 1 - Using qualitative evidence synthesis to inform guideline scope and develop qualitative findings statements. Health research policy and systems 17(1): 76.

Flemming, Kate, Booth, Andrew, Garside, Ruth et al. (2019) Qualitative evidence synthesis for complex interventions and guideline development: clarification of the purpose, designs and relevant methods. BMJ global health 4(suppl1): e000882.

Glenton C, Lewin S, Lawrie TA et al. (2019) Qualitative Evidence Synthesis (QES) for Guidelines: Paper 3 - Using qualitative evidence syntheses to develop implementation considerations and inform implementation processes. Health research policy and systems 17(1): 74.

Glenton, C; Lewin, S; Norris, SL (2016) Using evidence from qualitative research to develop WHO guidelines (Chapter 15). World Health Organization Handbook for Guideline Development. 2nd ed. Geneva: WHO.

Gould, Nick (2010) Integrating qualitative evidence in practice guideline development: Meeting the challenge of evidence-based practice for social work. Qualitative Social Work: Research and Practice 9(1): 93-109.

Hansen H.P.; Draborg E.; Kristensen F.B. (2011) Exploring qualitative research synthesis: The role of patient's perspectives in health policy design and decision-making. Patient 4(3): 143-152.

Knaapen, Loes, Colvin, Christopher J, Cowl, Jane et al. (2015). How to include qualitative research on patient views in guidelines GIN Public Toolkit: 28.

Kristensen FB & Sigmund H (ed.) (2007) Health Technology Assessment Handbook Copenhagen: Danish Centre for Health Technology Assessment, National Board of Health (https://www.sst.dk/~/media/ECAAC5AA1D6943BEAC96907E03023E22.ashx [accessed 08/09/19]).

Lewin, Simon and Glenton, Claire (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.

Lewin, Simon, Glenton, Claire, Lawrie, Theresa A et al. (2019) Qualitative Evidence Synthesis (QES) for Guidelines: Paper 2 - Using qualitative evidence synthesis findings to inform Evidence-to-Decision frameworks and recommendations. Health research policy and systems 17(1): 75.

National Institute for Health and Care Excellence (2018) Developing NICE guidelines: the manual.(https://www.nice.org.uk/process/pmg20/chapter/introduction-and-overview [accessed 08/09/19]).

Ring N, Ritchie K, Mandava L JR (2010) A guide to synthesising qualitative research for researchers undertaking health technology assessments and systematic reviews. Glasgow: Quality Improvement Scotland (NHS QIS).

Ring, Nicola; Jepson, Ruth; Ritchie, Karen (2011) Methods of synthesizing qualitative research studies for health technology assessment. International journal of technology assessment in health care 27(4): 384-90.

Swedish Agency for Health Technology Assessment and Assessment of Social Services (SBU) (2016) Evaluation and synthesis of studies using qualitative methods of analysis. Stockholm: Swedish Agency for Health Technology Assessment and Assessment of Social Services (SBU); 2016.(https://www.sbu.se/globalassets/ebm/metodbok/sbuhandbook_qualitativemethodsofanalysis.pdf [accessed 08/09/19]).

Tan, Toni P Y; Stokes, Tim; Shaw, Elizabeth J (2009) Use of qualitative research as evidence in the clinical guideline program of the National Institute for Health and Clinical Excellence. International journal of evidence-based healthcare 7(3): 169-72.

CHAPTER 3: THE CONDUCT AND REPORTING OF QUALITATIVE EVIDENCE SYNTHESES IN HEALTH AND SOCIAL CARE GUIDELINES: A CONTENT ANALYSIS

This chapter builds on the previous chapter by presenting a content analysis of QES undertaken by the UK National Institute of Health and Care Excellence. It aims to explore how closely the QES produced at NICE match with reporting standards developed for them.

While the previous chapter has summarised the recommendations in the methodological literature around using QES for guideline development, this chapter will assess what has actually been done by a major guideline development body in terms of conducting and using QES in the context of health and social care guideline production. The study focuses on the reporting of QES, partly because of a paucity of good quality tools for assessing the methodological limitations of QES (most are adaptations of quantitative tools), but mostly because the guideline committee discussions that will be further explored in future chapters are based on the evidence review that reports the QES and its results.

This chapter is published in BMC Medical Research Methodologies in 2022. Minor amends to the chapter have been made following examiner comments so there are slight differences from the published version.

Carmona, C., Baxter, S. & Carroll, C. The conduct and reporting of qualitative evidence syntheses in health and social care guidelines: a content analysis. *BMC Med Res Methodol* **22**, 267 (2022). https://doi.org/10.1186/s12874-022-01743-1

Abstract

Background: This paper is part of a broader investigation into the ways in which health and social care guideline producers are using QESs alongside more established methods of guideline development such as systematic reviews and meta-analyses of quantitative data. This study is a content analysis of QESs produced over a 5-year period by a leading provider of guidelines for the National Health Service in the UK (the National Institute for Health and Care Excellence) to explore how closely they match a reporting framework for QES.

Methods: Guidelines published or updated between Jan 2015 and Dec 2019 were identified via searches of the National Institute for Health and Care Excellence (NICE) website. These guidelines were searched to identify any QES conducted during the development of the guideline. Data relating to the compliance of these syntheses against a reporting framework for QES (ENTREQ) were extracted and compiled, and descriptive statistics used to provide an analysis of QES conduct, reporting and use by this major international guideline producer.

Results: QES contributed, in part, to 54 out of a total of 192 guidelines over the five-year period. Although methods for producing and reporting QES have changed substantially over the past decade, this study found that there has been little change in the number or quality of NICE QESs over time. The largest predictor of quality was the centre or team which undertook the synthesis. Analysis indicated that elements of review methods which were similar to those used in quantitative systematic reviews tended to be carried out well and mostly matched the criteria in the reporting framework, but review methods which were more specific to a QES tended to be carried out less well, with fewer examples of criteria in the reporting framework being achieved.

Conclusions: The study suggests that use, conduct and reporting of optimal QES methods requires development, as over time the quality of reporting of QES both overall, and by specific centres, has not improved in spite of clearer reporting frameworks and important methodological developments. Further staff training in QES methods may be helpful for reviewers who are more familiar with conventional forms of systematic review if the highest standards of QES are to be achieved. There seems potential for greater use of evidence from qualitative research during guideline development.

Key words: Qualitative evidence synthesis; Reporting frameworks; Guideline development

Introduction

Evidence-based health and social care guidelines (including clinical, public health and social care guidelines) are part of the landscape of evidence-based health and social care in many countries. These guidelines are normally based on one or more analyses of relevant evidence, often in the form of systematic reviews of effectiveness data and often interpreted by an expert committee.

Even though methods for synthesising qualitative research have been around for many years, interest in the use of qualitative evidence to inform the development of these guidelines has grown considerably over recent years. This is partly because of key developments such as more robust methods of synthesis, development of tools like GRADE-CERQual and better frameworks for reporting qualitative studies (100) and partly because qualitative data can answer particular types of questions better than quantitative data. Quantitative data are still key for questions of efficacy but are less able to answer questions relating to the effects of patient preference, feasibility, and acceptability on the broader effectiveness of a treatment or intervention. These questions are best answered by qualitative studies (101).

The World Health Organization (WHO) handbook (140) affirms that qualitative evidence should be used in the process of guideline development, and the Cochrane Qualitative and Implementation Methods Group are planning to publish a manual for qualitative evidence synthesis in February 2026. Other leading international guideline producers, such as the UK National Institute for Health and Care Excellence (NICE) are using qualitative evidence syntheses, both alone and as part of mixed-methods reviews, to present evidence to their guideline committees and this is supported by initiatives such as GRADE CERQual10 that have been developed with guideline committees specifically in mind. This surge of interest led Lewin and Glenton to declare “a new era” for qualitative research (130). A recent paper exploring how developers use qualitative evidence searched internationally for guidelines that used qualitative research and appraised their quality (141). The authors rated the guidelines using the AGREE II criteria, finding that most of the guidelines were of high-quality. However, the AGREE criteria are intended to assess the methodological quality of the guideline itself and the authors did not investigate the reporting of the evidence reviews that informed the guideline. They did not investigate the reporting of the evidence underpinning the guidelines.

A short paper published by Tan and colleagues in 2009 explored the use of qualitative evidence by NICE between 2002 (when NICE produced its first guidelines) and 2007 (109). The authors reported that almost 50% of NICE guidelines produced in that period made use of qualitative studies, although they did not report whether those are single qualitative studies or whether any qualitative evidence synthesis was undertaken. The paper noted a

growing trend by year in terms of the numbers of qualitative studies used in guidelines, rising from nine studies in 2003 to 41 in 2004, 60 in 2005 and 139 studies in 2006. The authors attributed the growth in the number of qualitative studies used to a combination of two factors. Firstly, a shift toward producing more guidelines on chronic conditions, where they argued that patient needs constituted an important part of the guideline, and secondly, that NICE's developing policy emphasis on patient and carer involvement led to more attention being paid to patient and carer perspectives.

They further noted that only five of the 22 guidelines which drew on qualitative research used (or documented) specific search strategies for qualitative literature over and above searches that were done for quantitative studies. Only four of the guidelines documented key methodological process details such as inclusion/exclusion criteria for qualitative studies.

This study also highlighted a gap in the reporting of the reviews - only half (11/22) of the guidelines reported how critical appraisal of qualitative studies was carried out, and only three of the 22 reported how data were synthesised.

The study concluded that "there is no consistency in how qualitative evidence is utilised in the development of NICE clinical guidelines. There are also clear training needs for NICE's guideline developers in terms of how best to identify, quality appraise and synthesise qualitative evidence" (p.172).

The work reported in this current paper updates the study by Tan and colleagues (109) by exploring whether methodological changes within NICE, or development in methodological standards for QES have led to a change in their use in NICE guidelines. It also builds on a review of methodological literature by the current authors (142). The study aims to examine all qualitative evidence syntheses used in guideline documents published between 2015 and the end of 2019 by a leading producer of guidelines for clinical, public health, and social care in the UK. NICE was chosen as an appropriate exemplar because of its international reputation as a leading guideline producer. The study aimed to explore where and how QES are used in the development of health and social care guidelines, and how the methodologies used compare with international standards of good practice.

Method

The study used a content analysis method to analyse textual data (143). Berelson (144) described content analysis as "a research technique for the objective, systematic and quantitative description of the manifest content of communication" (p. 18). Content analysis incorporates both quantitative approaches that convert the textual data to numerical data, for example by counting occurrences of the content of interest, and also more qualitative approaches that analyse the way that the content of interest in presented or discussed. The process followed in this study was based on the method outlined by Bengtsson (145) (see Table 3).

Table 3: Summary of Bengtsson method for content analysis

Stage	Tasks	How was this operationalised?
Planning	• Aim • Sample & unit of analysis • Data collection Method of analysis • Practical implications	• Aim – to better understand variation in the reporting of QES used in NICE guidelines • Sample – NICE guidelines published or updated 2015 – 2019 • Unit of analysis – A single QES was the unit of analysis rather than the guideline as a whole since some guidelines have multiple associated QES • Data collection/analysis – see boxes below • Practical implications – understanding where QES in the sample do not meet the criteria set out by ENTREQ is a useful indicator of reporting quality.
Data collection	• Collect data and transform to analysable text	• Overall set of eligible guidelines identified using functionality on NICE website. • Manual sifting of reviews undertaken for guidelines to identify QES. • QES downloaded as pdf documents for analysis.
Data analysing	• Categorisation • Compilation	• ENTREQ reporting criteria used as framework for categorisation with

		each element assessed as 'met' or 'not met' • Compiled in tabular form in spreadsheet.
Reporting	• Creating a report/presentation of the result.	The results are presented in this paper

Source documents

In order to compare recent NICE guidelines with the sample included by Tan et al (109), and to reflect current practice, we scrutinised guidelines from a 5-year period (the beginning of 2015 until the end of 2019).

Using inbuilt functionality on the NICE website, a search was conducted for guidelines published between January 2015 and December 2019. This search encompassed the three types of evidence-based guideline produced by the guideline development centres at NICE, classified on the website as 'public health', 'social care' or 'clinical'. It does not include guidelines where the method of development differed, that is, antimicrobial guidelines, cancer service guidelines, COVID-19 guidelines, and medicines practice guidelines (less than 40 guidelines in total). The resulting list of guidelines was copied to the clipboard (using the website functionality) and pasted into an excel spreadsheet (Microsoft Office Professional Plus 2019).

For each included guideline, the individual evidence reviews (systematic reviews and qualitative evidence syntheses) were explored using the 'evidence' tab on the guideline webpage.

Each evidence review was examined to evaluate whether or not a qualitative evidence synthesis (defined as two or more qualitative studies combined together to answer the same review question) had been undertaken by the technical team (or a contractor) responsible for the development of the guideline. Evidence reviews that did not report the use of qualitative evidence synthesis (or mixed-methods synthesis with a qualitative component) were excluded from the sample. Any qualitative reviews and mixed-methods

reviews identified were downloaded and saved. These formed the sample for the content analysis.

Data Collection

Included QES were copied to a new excel spreadsheet and rationalised so that the unit of analysis was the qualitative evidence synthesis rather than the guideline (some guidelines were supported by multiple qualitative evidence syntheses). The coding framework (described below) was added to the spreadsheet to create a data extraction tool.

The coding framework used was intended to provide two sets of data – descriptive data and content data.

Descriptive Data

This included key data from the QES – guideline number, year of publication, author (by guideline-producing centre rather than individual authors) and number of qualitative studies included in the analysis. The use of GRADE-CERQual (110) to assess the confidence was also noted.

Content Data

The criteria set by ENTREQ (64) are the most commonly used reporting framework for QES, and therefore this framework was selected as a useful one for examining the content of the QES included in this study – see Table 4 and appendix F. There are alternative reporting standards for specific types of QES, for example the eMERGe Reporting Guidance for meta-ethnography (146), but since NICE has not produced any of these types of QES they were not used in this analysis.

Table 4: Summary of ENTREQ criteria

- Aim	- Appraisal items
- Synthesis methodology	- Appraisal process
- Approach to searching	- Appraisal results
- Inclusion criteria	- Data extraction
- Data sources	- Software
- Electronic Search strategy	- Number of reviewers

- Study screening methods	- Coding
- Study characteristics	- Study comparison
- Study selection results	- Derivation of themes
- Rationale for appraisal	- Quotations
	- Synthesis output

Data Analysis

Each of the QES was read and descriptive data and content data were coded into an excel spreadsheet according to the framework described above and in appendix F. Coding was binary (0 or 1) and indicated whether the QES reported on the criterion in the reporting framework or not. For example, did the QES report its aim? Did it report the synthesis methodology it is underpinned by? This approach did not allow for any judgment about the adequacy of each reporting criterion, only whether it was present or not. This approach was taken to allow for analysis of coding.

Resulting data are presented predominantly as descriptive statistics to show trends, consistencies, and inconsistencies in the data. Data were imported into the R program (147), using the 'tidyverse' package (148) to manage the data and the 'ggplot2' package (149) (also part of the tidyverse) for data visualisation. The R code used to generate the figures can be found in appendix G.

Results

Number and Size of QES Undertaken

Between January 2015 and December 2019, NICE published 192 clinical, public health and social care guidelines. The website categorises the breakdown of these guidelines as 156 clinical, 30 public health and 48 social care guidelines, however this includes some guidelines listed in more than one category, hence the discrepancy in numbers. For the purposes of this analysis, pragmatic decisions were made about the main topic area of a guideline to assign each guideline to a single category, resulting in a breakdown of 143 clinically focussed guidelines, 25 public health focussed guidelines, and 24 social care focussed guidelines. Each of these guidelines is based on multiple sources of evidence – most often systematic reviews of quantitative evidence, but also prognostic and diagnostic

reviews (of the predictive or diagnostic accuracy of tests or indicators), epidemiological studies (of prevalence and incidence) and, more rarely, qualitative evidence syntheses. The total number of reviews (both quantitative and qualitative) conducted for a guideline can range from one review for an update of a single clinical question to around 40 reviews for a large guideline with multiple questions. The reviews are conducted by expert review teams who present them to the guideline committee. The committee undertake a structured discussion (although not using a formal Evidence-to-Decision framework) of the evidence contained in the reviews (and their confidence in that evidence if GRADE-CERQual was used), alongside any other evidence, and contextualise it using their expertise and experience of the UK health and social care system to make guideline recommendations. When a guideline is published, all of the evidence considered by the committee is also published alongside the guideline.

Of the 192 guidelines referred to above, 54 guidelines (28%) had one or more QES as part of their evidence base (qualitative evidence syntheses defined as a synthesis of more than one qualitative study). Overall, out of a total of approximately[b] 1,500 reviews/research questions, 90 were QES (approx. 6%).

Of the 54 guidelines with one or more QES, 36 (out of a total of 143 [25%]) were clinically focussed, 13 (out of 25 [52%]) were public health focussed, 5 (out of 24 [21%]) were social care focussed. This shows that social care and clinically focussed guidelines are roughly half as likely to use qualitative evidence synthesis as public health focussed guidelines (see **Table 5**).

Table 5: Prevalence of QES by guideline type

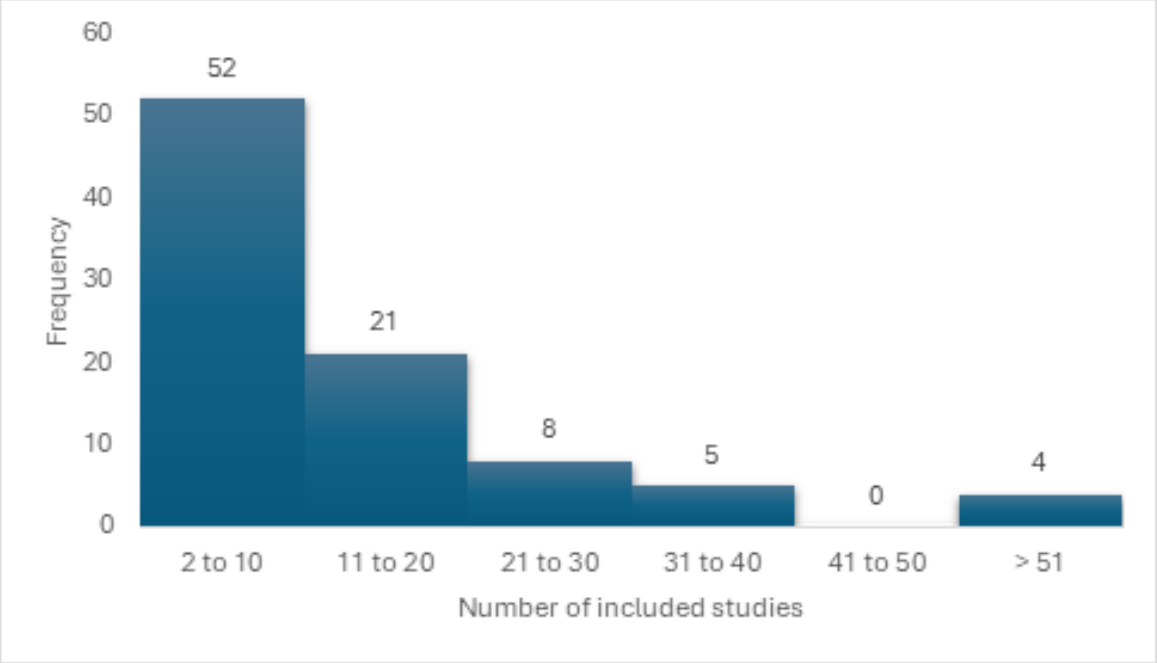
	Guidelines published in study period (n=192)	**Number with QES (n= 54)**	**Percentage of total**
Clinical	143	36	25%
Public health	25	13	52%
Social care	24	5	21%

[b] It is not possible to accurately count the number of review questions due to changes in the way that these are reported.

The number of QES used per included guideline ranges from 1 to 6 (mean = 1.67 per guideline that contains a QES, less than 0.4 QES per guideline published between Jan 2015 and Dec 2019).

In terms of the number of included papers in the QES, there was a large amount of variation. The largest QES contained 69 papers, the smallest QES contained two papers. Distribution of QES by the number of included papers is shown in Figure 3. Reasons for the variation were not explored as part of this analysis but may be related to the size of the evidence base, or to the formulation of the review protocol.

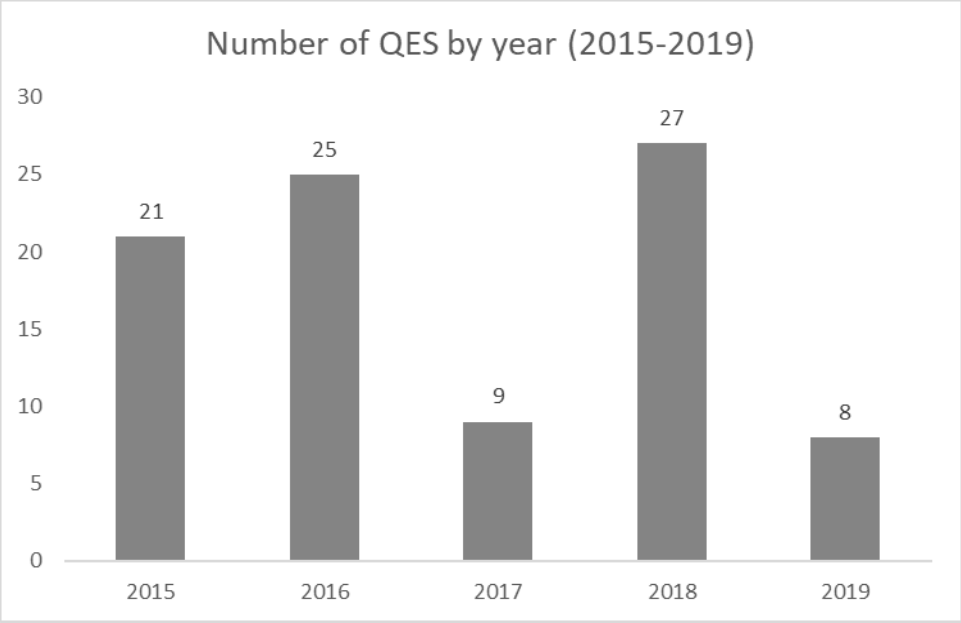
Figure 3: Frequency of QES by number of included papers



Overall, 65% (58 out of 90) of QES had less than 12 papers included, with a mode of four and a median of ten papers. The four QES with more than 42 papers were from two guidelines (150, 151) and in both cases a single set of included papers was identified through searching and sifting and the data were extracted from the single set of papers to develop two QES with different review questions.

Figure 4 shows the number of QES conducted by year for the period 2015 – 2019. The graph does not indicate any meaningful trend toward producing more QES in spite of the growth in acceptability of QES in evidence-based health and social care, and the development of more rigorous methods (see methodological review). The large variations in 2017 and 2019 might be at least partly explained by the lifecycle of a guideline. In most cases guidelines take longer than a year to develop and publish. The number of guidelines published per year is somewhat variable, depending on the length of the guidelines' development – guidelines with more review questions, usually addressed sequentially, tend to have longer development times. There is no evidence found by this analysis that would indicate why 2017 and 2019 were years when fewer QES were published.

Figure 4: Number of QES published by year (2015-2019)



Purpose of QES undertaken

There are a range of QES methodologies which vary widely on the epistemological spectrum, and in level of complexity, from aggregative approaches to more configurative/interpretive approaches. QES undertaken for NICE guidelines all use simpler descriptive or aggregative approaches. These syntheses can be used to address a range of issues that concern people's (both patients and healthcare professionals) views, beliefs and lived experiences. While quantitative evidence is best for addressing questions of efficacy

(does treatment A have an effect on condition B?), qualitative evidence can be useful to bridge the gap between efficacy and real-life effectiveness, for example understanding why people do not take their medicines as prescribed, how the medicines impact their lives and how things could be improved. In spite of this, guidelines produced by NICE in the period 2015-2019 seem to address a much more limited range of question types using QES. Almost half of the QES undertaken answer one of two types of question:

- What are the barriers and/or facilitators to……?
- What are the information (and support) needs of ……?

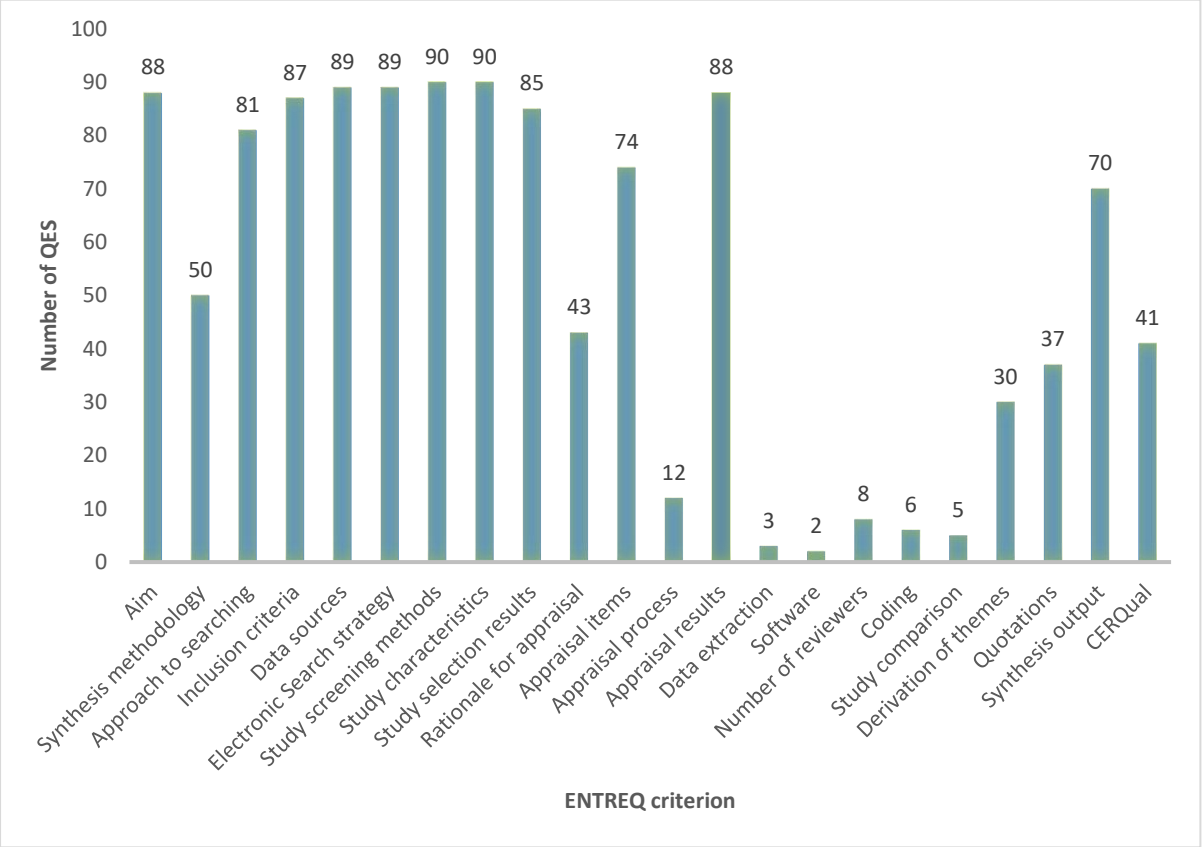
Many of the remaining questions deal with similar question types, often about support and care needs. This may indicate a limited understanding in the NICE guideline development centres of the potential remit of QES and their flexibility with regards to issues such as service configuration, professional support etc. Other types of QES do include occasional innovative questions, for example one QES for guideline NG77 - management of cataracts in adults (152) was employed to explore how lens implant errors happen through qualitative analysis of physician reports and case studies.

Quality of Reporting

The 90 QES published by NICE between Jan 2015 and Dec 2019 were assessed against the ENTREQ reporting criteria as described in Table 4 (above) and in more detail appendix F.

Analysis of number of guidelines meeting each of the ENTREQ criteria is shown in Figure 5 with an additional column to indicate whether the QES used GRADE-CERQual to assess confidence in the qualitative findings.

Figure 5: Number of QES meeting each ENTREQ reporting criterion (out of a total of 90)



ENTREQ criteria relating to setting out the aim of the review and to the systematic searching and sifting of studies to generate a pool of included studies was generally done well and described adequately in the included QES. Almost all of the QES provided a structured review question or stated aim and a full search strategy and PRISMA flowchart to show the flow of studies. In line with good practice, they provided reasons for exclusion for papers excluded at full-text evaluation. The exception to this was the synthesis methodology criterion (described by the ENTREQ statement as "Identify the synthesis methodology or theoretical framework which underpins the synthesis and describe the rationale for choice of methodology"). Many QES (40/90) were given a 0 mark on this criterion because either they only provided a brief sentence or statement to describe the methods of data synthesis used, for example "We undertook thematic synthesis", with no methodological detail, or simply provided inadequate descriptions of methodology, often not specifying an approach to synthesis at all.

Derivation of themes (described by the ENTREQ statement as “Explain whether the process of deriving the themes or constructs was inductive or deductive”) was demonstrated in a third of QES, and these were mostly undertaken by a particular guideline developer who present a ‘theme map’ as a standard part of their QES.

In 70 of the reviews, synthesis output (described by the ENTREQ statement as “Present rich, compelling and useful results that go beyond a summary of the primary studies”) was reported. This was mostly in the form of NICE evidence statements, although some evidence statements made no attempt at synthesis and simply listed the themes identified by individual studies. Some QES used a Cochrane style ‘Summary of qualitative findings’ table to present synthesised themes and sub-themes along with their CERQual confidence rating. Other than that, CERQual was not often used. This does not seem to be dependent on the age of the review (as might be expected given the introduction of CERQual in 2015) but seems to depend more on the guideline developer.

Variation Over Time

It might be expected that adherence to reporting frameworks improves over time as methods for undertaking QES become more robust and more widely known. It might also be expected that guideline developers would develop their methods for QES (and train their staff in those methods), and that more recent iterations of the NICE guideline methods manual might give clearer direction on its expectations from QES.

Figure 6 explores how well QES from different centres match with criteria in the ENTREQ reporting framework over time. For years where a centre produced more than 1 QES, the mean of the number of criteria in the framework (out of 21) for the QES produced in that year is used. It is important to note that using a mean number of reporting criteria is somewhat arbitrary since it requires making a generalisation that each of the 21 criteria in the framework is of equal importance to the reporting of a QES.

Data suggest that in fact there is little variation over time, but that the main determinant of the number of ENTREQ criteria reported is the guideline developer who authored the review. Of the two guideline developers who authored the majority of the QES in the past 5 years, one reasonably consistently reports around 11-13 criteria (Centre 7), whereas the other performs better in 2016 and 2017, but drops to a similar level in 2018 and 2019

(Centre 6). It is unclear what may drive the drop. Two possible confounding factors are the publication of the new NICE methods manual in 2018 (105), or simply a change in staff or senior staff from someone more familiar with QES to someone less familiar.

To further explore this, data were plotted to calculate the median number of ENTREQ criteria reported over all years (2015-2019) by guideline developer. Figure 7 presents this data along with the associated point values for each QES.

Figure 6: ENTREQ criteria (out of a maximum of 21) reported by year and authoring centre

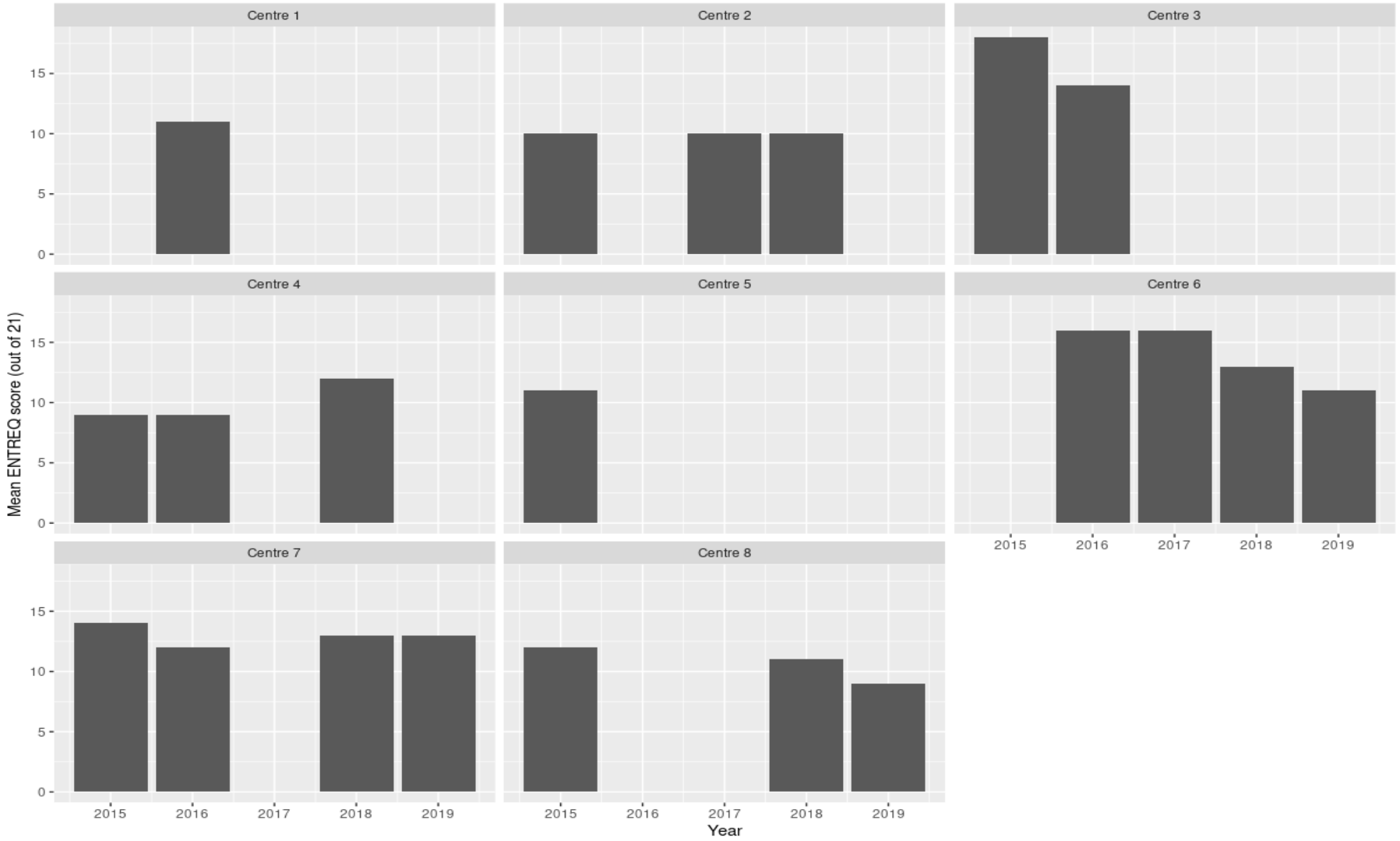
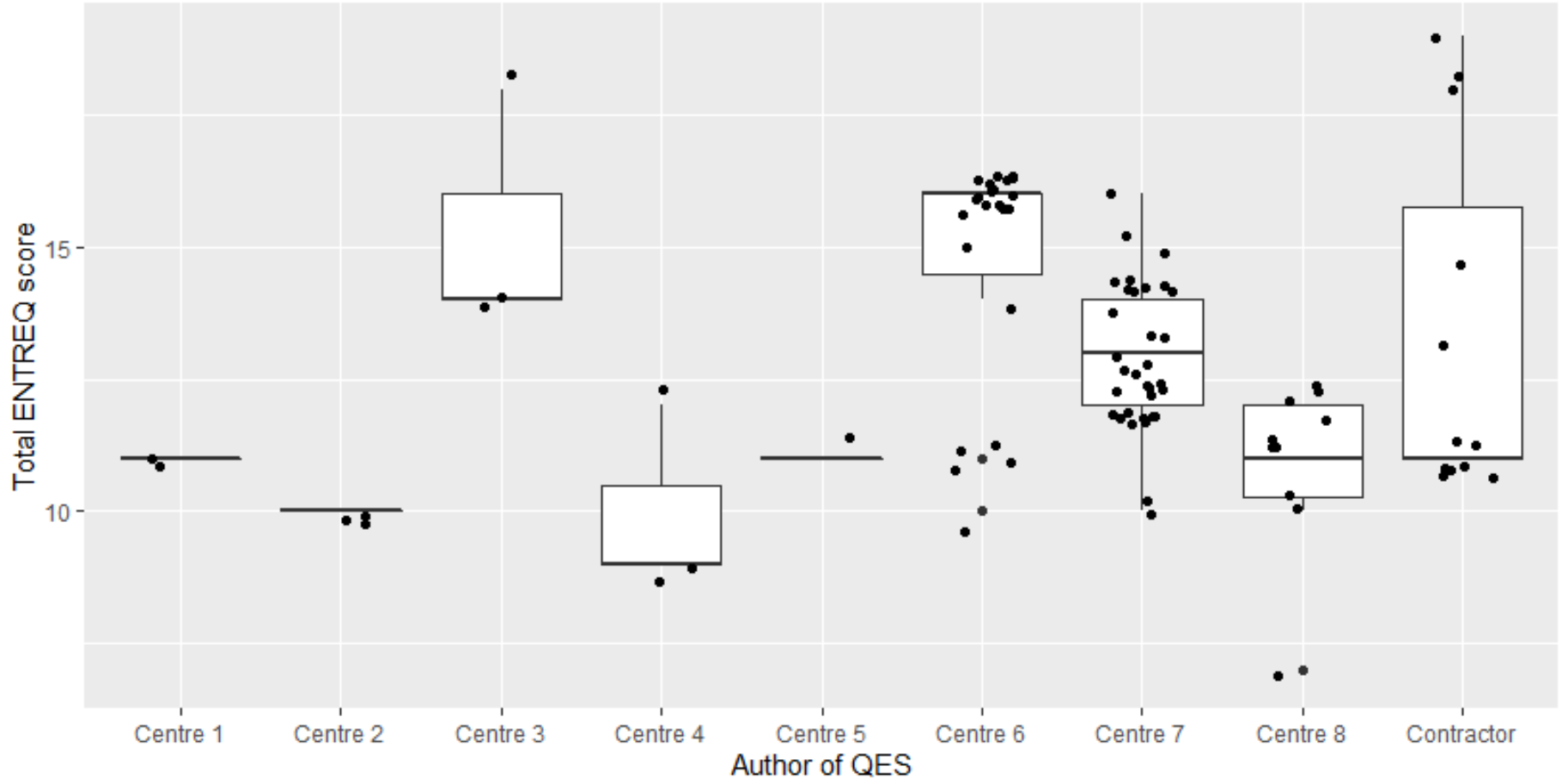


Figure 7: Median number of criteria in the ENTREQ framework met

(dots represent individual QES)



The data in Figure 7 broadly support the hypothesis that the different producers of QES account for most of the variation in the number of criteria reported on in the reporting framework. Centres that do less well tend to have only produced 2 or 3 QES over the 5 years period and therefore staff are likely to have been less familiar with QES methods having done them rarely. The Centre 8 team do not appear to fit this pattern. Their QES perform poorly against the ENTREQ framework, however the team have produced 11 QES in the 5-year timeframe, including the lowest scoring and second lowest scoring.

The widest variation in meeting the criteria in the framework is seen in the contractor group, but this is to be expected since it is a heterogeneous group comprised of various organisations and academic groups. Since these QES were contracted out, it is reasonable that the highest ranking QES are in this group since competitive tendering would lead to these syntheses being undertaken by specialist teams familiar with QES.

Centres 6 and 7 are the most prolific producers of QES, with centre seven demonstrating a wide range of reporting quality across their QES. Centre 6 reporting quality appears to be dichotomous with a cluster of QES scoring 10 or 11, and a larger cluster scoring 15 or 16. It is unclear what the cause of this dichotomy might be.

Discussion

Number and Size of QES Undertaken

The number of QES undertaken by NICE (including its contractors) over the 5-year period up to the end of 2019 formed a fraction of the total number of reviews undertaken in the period. Although it is difficult to ascertain why this is the case, there are plausible explanations that can at least partially explain this lack of attention to the qualitative evidence.

The majority of the guidelines produced in the period were clinical guidelines (143 out of 192), and clinical guidelines are most often about the relative efficacy of different treatment modalities. In questions of efficacy, the gold standard is the randomised controlled trial, or a systematic review of randomised controlled trials. Although QES could be used to bridge the efficacy – effectiveness gap (that is, the difference between the biological or medicinal effect of the medicine itself on the body and its observed effectiveness in a particular

population) by addressing issues such as acceptability of the treatment, compliance with regimes, attitudes towards the medicine etc., the reality is that in the majority of cases there is unlikely to be published qualitative evidence that could be synthesised that directly addresses the efficacy question. For example, while there might be substantial research into peoples lived experiences of particular illnesses, there is less likely to be evidence on people's experiences of undergoing treatment A specifically. The most obvious exception to this is in long-term conditions, or conditions where there is a notable impact on quality of life, where there is potentially substantial qualitative research – for example, cancer care or kidney dialysis. There is also a growing recognition within producers of clinical guidelines of the importance of qualitative evidence as a tool in implementation research because they "generate opportunities to examine complexity and include a diversity of perspectives" (153).

Arguably, QES could be more routinely useful in public health and social care topics where interventions tend to be more interpersonal or sociopsychological than biological and evaluations of views, perceptions and lived experiences (traditionally the domain of qualitative research) are more likely to be qualitative than in clinical medicine.

The line of argument about the likely availability of qualitative data is to a large extent borne out by the size of the QES that were carried out. With a modal number of four papers per QES they are, on average, relatively small. Themes from QES that contain so few studies may not score highly in a CERQual assessment (they are likely to be downgraded for adequacy unless the data from the studies is very rich), and this may restrict their usefulness as part of a decision-making process. Of the four large (>50 papers) QES, two were part of the workplace health guideline16, a non-clinical, public health guideline, and two were related to the attention deficit hyperactivity disorder: diagnosis and management guideline (151), which fits the model of a long-term condition with a notable impact on quality of life.

It is also plausible that the lack of relevant studies identified for most of the QES was due to either inappropriate research questions, or insufficient searching. Technical staff and information specialists producing QES within NICE are usually quantitative systematic reviewers and have little training in searching for or assessing qualitative evidence. Added to this, qualitative studies are notoriously poorly indexed in databases, qualitative study filters

are still quite primitive in comparison to quantitative ones, and qualitative literature searches are often quite specific (as opposed to sensitive) to limit the large amounts of irrelevant papers that need to be excluded during the sifting process (154).

The numbers of QES published per year does not appear to have the incremental increase that would be expected given the development of methods for QES over the 5 years in question, however this could be simply because the time period is too short to demonstrate any trend. It is also likely due to the varying patterns of NICE guideline publication. NICE guidelines take varying amounts of time to complete depending on a variety of factors, so there is not a consistent background rate of guideline publication against which the numbers of QES can easily be measured. The Tan paper (109) however, reports that almost 50% of guidelines published in 2002 – 2007 'made use of qualitative studies' (this is a slightly different measure to 'undertaking a QES' – the inclusion criterion for the current study. See below). During 2015 – 2019 that number was 28%, so a more detailed examination of the numbers over the lifetime of NICE could potentially reveal a year on year decrease in the number of guidelines using QES. A caveat here is that the Tan paper (109) refers to 'making use of qualitative studies' but does not define this. There are guidelines from that period that report single or small numbers of qualitative studies but do not make any attempt at synthesis and therefore would not be considered for this study. The current content analysis only counted syntheses of two or more qualitative studies and did not count incidental use of single qualitative studies. This change from counting single studies to only counting syntheses of two or more studies is likely to account for a good deal of the discrepancy.

Purpose of QES Undertaken

Almost half of the QES undertaken in 2015 – 2019 were carried out to address generic questions about barriers and facilitators to accessing a service or treatment, or about information needs relating to a condition. A substantial number of the remainder were about care and support needs of people with a specific condition. There seems in general little appetite to address more creative questions through QES even though the NICE manual (105) gives a broader list of examples than this including:

- What elements of care on the general ward are viewed as important by patients following their discharge from critical care areas?

- How does culture affect the need for and content of information and support for bottle or breastfeeding?
- What are the perceived risks and benefits of immunisation among parents, carers, or young people? Is there a difference in perceived benefits and risks between groups whose children are partially immunised and those who have not been immunised?
- What information and support should be offered to children with atopic eczema and their families and carers?
- What are the views and experiences of health, social care, and other practitioners about home-based intermediate care?

Occasional forays are made into more novel uses of QES. For example, in the "Cataracts in adults: management" guideline (152), a QES was undertaken to inform recommendations on wrong lens implant errors, specifically the questions "What are the procedural causes of wrong lens implant errors?" and "What strategies should be adopted to reduce the risk of wrong lens implant errors?".

An avenue that does not seem to have been routinely explored by NICE is the use of QES as contextual grounding for guidelines. For example, a guideline about diabetes might usefully be underpinned as a whole by a QES that explored people's experiences of living with, or caring for, people with diabetes, even though qualitative data to inform a QES about specific question within the guideline might not be available, the context would enable a guideline committee to frame their recommendation-making in terms of people's lived experience of the condition.

Quality of Reporting

It is clear from Figure 5 that there is good consistency within the ENTREQ criteria as to whether it is done well or poorly in NICE QES. Most criteria are either reported on by over 80 (out of 90) or by less than 45 QES. Very few criteria fall between these brackets.

Closer examination of the reporting criteria reveals that the criteria in the framework where the number of QES reporting the criterion are very high are all criteria that duplicate steps in quantitative systematic reviews and are therefore familiar to staff who are predominantly quantitative systematic reviewers. ENTREQ criteria relating to documenting the searching and sifting process, and to the creations of evidence tables of study characteristics are

invariably done well, as is the presentation of the results of the methodological critical appraisal of the papers. Almost all of the criteria that duplicate steps in the quantitative systematic review process were reported in the QES (85 or more of the 90 QES).

Steps that are unique to QES, or where QES methods differ from quantitative systematic review methods, fare less well, and this is particularly the case with the criteria in the framework that require specific skills in methods for QES: data extraction, coding, use of software, and study comparison all fare poorly with less than 10% of the included QES reporting how (or if) they undertook these steps. Description of methods of qualitative synthesis also fared poorly with only around half of the QES reporting a synthesis approach in any detail.

Variation Over Time and Centre Undertaking QES

The data presented here for different guideline-producing centres are, at best, only indicative data. The picture they present of static guideline-producing centres is potentially a misleading one. In the period under scrutiny (2015-2019), major changes were made to the way in which NICE contracts out work for guideline production. In the early stages of this time period, NICE had contracts with several external collaborating centres, mostly associated with academic units, and additionally an internal clinical guidelines team and a public health team. The external teams were responsible for specific areas of guideline production (for example, the National Collaborating Centre for Mental Health, or the National Collaborating Centre for Women's and Children's Health). The collaborating centres were replaced with two generic bodies, the National Guidelines Alliance and the National Guidelines Centre. These two bodies absorbed the functions, and in many cases the staff, of the collaborating centres. It is likely that the changing membership of review teams over that time has had an impact on the systematic review and QES processes that underpin the guidelines (155).

In spite of this, there seem to be two general trends in the data contained in Figure 6 and Figure 7 that are important for this analysis. Firstly, that over time the completeness of reporting of QES both overall, and by specific centres, has not improved in spite of clearer reporting frameworks and important methodological developments in QES. Secondly, the quality of reporting seems (in most cases) to be related to the centre producing the QES,

with clear clusters of reviews of similar quality within centres. The exceptions, as discussed above, are the generic 'contractor' category and the public health team.

Along with its international peers, including Cochrane and the WHO, NICE is developing methods for the use of QES in producing health guidelines (105). To date this seems only to have been through relatively small numbers of QES, and only to address a very limited number of questions, primarily those about barriers and facilitators to service use and about people's information and support needs when diagnosed with, or living with, a health condition. There is a potential to better understand the range of questions which qualitative evidence might be able to shed light on, and this in turn might make them more common as part of guideline production.

The focus of health guideline-producing bodies on the use of systematic reviews of quantitative evidence and the relatively small amount of QES means that there was no noticeable improvement over time in the quality of QES produced. QES that were not produced by contractors who specialise in qualitative methods often lacked transparent reporting of those aspects of the qualitative evidence synthesis that differ from the stages of a quantitative systematic review.

The clearest factor in the quality of a QES seems to have been the team that undertook it. Teams which produced well-reported QES seemed to do so consistently, and we can speculate that this may because they have staff with a particular interest or skill set in this area.

Limitations

While we believe that the findings are robust, we acknowledge that the way that reviews are reported by NICE changed several times during the 5-year period under consideration. At various times multiple questions could be subsumed into single reviews or split across different review questions. This means that accurate counting becomes difficult, and some numbers are a near approximation based on counting and pragmatic decisions. Where numbers are uncertain this is reported.

The ENTREQ framework was not intended to be used for 'scoring' QES, and arguably not all ENTREQ reporting domains are equal in importance, nor was it designed as a formal

reporting standard - it is a general statement containing 21 items or criteria that can be broadly applied to common types of QES methodologies. As a framework, it is not well suited for more complex methodologies, however it is useful for simpler descriptive/aggregative methods as used in the QES described here.

The main purpose of this analysis was to better understand the quality of reporting of QES rather than why QES were or were not undertaken for specific guidelines. QES are relevant to a very specific range of research questions, and not all NICE guidelines would have benefitted from a QES. Further research would need to be undertaken to establish whether QES had been used appropriately in guideline development.

As with any documentary appraisal, it is unclear whether issues identified in this paper are due to the lack transparent reporting of the qualitative evidence syntheses or whether they relate to the conduct of the reviews themselves, or just to the reporting of them.

Conclusions

This study analysed guidance containing a QES from one UK organisation over a 5-year period. Analysis found that the use, conduct and reporting of optimal QES methods requires development, as over time the quality of reporting of QES both overall, and by specific centres, has not improved in spite of clearer reporting frameworks and important methodological developments. Further staff training in QES methods may be helpful for reviewers who are more familiar with conventional forms of systematic review if the highest standards of QES are to be achieved. There seems potential for greater use of evidence from qualitative research during guideline development. The studies reported in Chapters 2 and 3, provide insight into the methods used to develop QES in health and social care guideline-producing organisations (Chapter 2) and what sort of QES has been used by NICE guideline committees in particular (Chapter 3). However, the literature is largely silent on the fate of that evidence at the hands of a guideline committee or panel. The next study reported in this thesis attempts to begin to remedy that situation by focussing in on the way that guideline committees understand and use the QES that are put in front of them, in the context of the wider (quantitative) evidence base, and their own beliefs and opinions about the value and utility of QES.

CHAPTER 4: METHODS FOR THE PRIMARY QUALITATIVE STUDY OF COMMITTEE MEMBER AND TECHNICAL STAFF VIEWS AND OPINIONS ON THE USEFULNESS OF QES IN GUIDELINE DEVELOPMENT

Introduction

This chapter and the following chapter address the overall question for 'phase 2' of the PhD; to explore the views and perceptions of technical staff, committee experts and committee lay-members regarding how a QES contributes to committee discussions, and to the process of making recommendations. It explores the gaps highlighted in Chapter 2 relating to establishing whether guideline-producing committees find QES useful to their deliberations, whether they could be done or presented differently to make them more useful and, perhaps most importantly (the gap identified in Chapter 3), how committees use QES to inform their decision making alongside quantitative systematic reviews of effectiveness and colloquial evidence (such as expert testimony and the committee members own expertise and experience). It further aims to explore committee members and evidence synthesis team members understandings of QES and how it is reported (textually) and presented (verbally) to them to reflect on the importance (or otherwise) of technical expertise in reviewing teams and among committee members as proposed in Chapter 3.

This chapter describes the methodological approach taken to conduct a primary qualitative study that was designed to elicit the beliefs and opinions of members of NICE guideline committees about the usefulness of QES during the guideline production process. It builds on and complements 'phase 1' of this research - the two previous studies (the systematic review of the methodological literature and the content analysis of QES adherence to reporting standards) which focussed on the content and methods of producing QES in the context of guideline development. It does this by exploring whether QES, as they are presented to NICE committees, help the committee decision making processes and drive the formulation of guideline recommendations, and what factors help or hinder this process.

Reporting Criteria

In spite of recent concerns about its development (156) and its legitimacy as a reporting tool (157), the COREQ reporting criteria are the most frequently used reporting criteria, and therefore this study followed those criteria. A completed COREQ checklist for this qualitative study can be found in appendix H.

Context and setting for the qualitative study

The UK's National Institute for Health and Care Excellence (NICE) is one of the world's leading producers of evidence-based guidance in health and social care (158). Guidelines are primarily informed by systematic reviews and meta-analyses of randomised controlled trials (RCTs) and economic analyses that are interpreted by guideline committees formed of experts, clinicians, and lay-members. The Institute occasionally considers evidence from observational studies, and rarely, from qualitative studies. This tends to be in exceptional circumstances when not enough RCT evidence is available to inform a decision, or when a question cannot be answered with experimental evidence, for example questions about information needs or barriers to service access (see Chapter 3 for a breakdown of the kinds of questions addressed by QES at NICE).

NICE was selected as an appropriate setting for this study because:

- NICE is considered to be a world leader in evidence-based health guideline production and findings could be applicable to other organisations around the world which produce healthcare guidance.
- I am an employee of NICE and have an insider understanding of NICE processes and was therefore well placed to access NICE committees for data collection, although it is recognised that as an employee, I would have particular viewpoints/perceptions. This is considered in the ethics section below.

In line with most international guideline-producing bodies, NICE makes use of guideline committees (often known as guideline panels outside of NICE) to make recommendations for guidelines. The committees are composed of an independent Chair, relevant experts (often clinicians or other healthcare professionals) representing the multidisciplinary team relevant to the topic of the guideline, and at least two lay-members (representatives of

patient/service user voice organisations, or individuals with lived experience of a condition about which the guidance is produced, or of caring for someone with that condition). Further information about NICE's approach to guideline production is available via their website (159), and methodological and process detail in the NICE Guidelines Manual (30).

Technical teams of systematic reviewers at NICE produce evidence syntheses based on review protocols agreed by the committee to address the issues identified as relevant to the guideline. Normally these are systematic reviews of RCTs, or of observational evidence. Occasionally these reviews are of qualitative evidence.

The evidence reviews are sent to committee members in advance of a guideline committee meeting. These meetings are normally a day in length and normally (since COVID-19) conducted virtually. At this meeting committee members are presented, by NICE technical staff, with a summary of the key evidence contained in the evidence reviews, along with any health economic evidence or modelling that has been undertaken. The committee discuss the evidence in the context of their own experience and practice, and what they know of the broader UK context to produce recommendations for the guideline.

The recommendations take the form of statements aimed at healthcare professionals (HCPs) to guide their practice, either as 'strong' recommendations that tell HCPs how they **should** act in a particular situation, or 'weak' recommendations that HCPs might **consider** in a particular circumstance (see Figure 8 for an example from NICE's adrenal insufficiency guideline (160)). Recommendations are based on a GRADE assessment of the confidence in the evidence (or a GRADE-CERQual assessment for QES), and on the strength of the evidence (for example, the size of the pooled effect estimate if a meta-analysis was undertaken), However, it is also within the committee's power to make consensus recommendations that can be based on expert testimony, non-systematic review sources of evidence or, in the absence of any other evidence, based on committee consensus.

Figure 8: Example strong and weak recommendations from a NICE guideline

Example of a strong recommendation

1.4.6 Admit the person to hospital during periods of physiological stress if they are unable to absorb oral glucocorticoids, for example, during prolonged diarrhoea and vomiting.

Example of a weak recommendation

1.3.5 For people with primary adrenal insufficiency and persistent hyponatraemia despite having the maximum dose of fludrocortisone, consider sodium chloride supplementation according to specialist endocrinology advice.

(recommendations taken from NICE guidance on **Adrenal insufficiency: identification and management)**

After the meeting, a narrative description of the committee's deliberations is included in the evidence review document to demonstrate the links between the evidence presented and the recommendations the committee made. Once completed, draft guidelines undergo public consultation and may undergo revision before publication.

Research question

The overall research question for this thesis is:

- What is the role of qualitative evidence syntheses in the development of health (including public health) and social care guideline development, and how might its role be optimised?

Previous phases of work have explored the existing literature on how QES should be undertaken to inform guideline committee decisions (Chapter 2 – systematic review of the methodological literature) and how NICE has reported QES undertaken previously for its guidelines by comparing the content of the QES to the ENTREQ reporting standard (Chapter 3 – Content analysis of NICE QES). The primary qualitative research element of the PhD builds on this previous work by exploring in-depth the perceptions of key individuals involved in guideline development. The qualitative study aimed to specifically address the following research questions:

- How do guideline committees at NICE use findings from QES to inform recommendation-making?
- What are the views and perceptions of technical staff, committee experts and committee lay-members regarding how a QES contributes to committee discussions, and to the process of making recommendations?
- What might be learned about best practice?

Epistemological underpinning of the qualitative study

The research was grounded in the framework of evidence-based health care (EBHC). Previously known as evidence-based medicine (EBM), EBHC is a clinical, public health, and social care approach that emphasises the use of the best available evidence to inform healthcare decisions (12). While EBHC has been influential in shaping clinical practice, there is a growing recognition of the complexities involved in translating evidence into guidelines (17). This study will explore these complexities by examining how guideline committees currently and in the future might best navigate the challenges of interpreting and integrating evidence from QES into their recommendations.

In the realm of EBHC, adopting a critical realist perspective offers a robust framework for understanding the complexities of healthcare interventions and guidelines. Critical realism, a philosophy developed by Roy Bhaskar, bridges the gap between positivism and constructivism by acknowledging that reality exists independently of our perceptions, yet our understanding of it is always mediated by social and subjective factors (161).

Key Principles of Critical Realism

1. **Ontological Realism**: Critical realism posits that there is a reality independent of human thoughts and beliefs. This reality includes both physical and social structures that influence health outcomes. For example, the biological mechanisms of a disease exist regardless of our understanding or awareness of them.
2. **Epistemic Relativism**: While reality is independent, our knowledge of it is always partial and fallible. This means that scientific knowledge is socially constructed and influenced by the context in which it is produced. In EBHC, this principle encourages researchers to critically examine how evidence is generated and interpreted.

3. **Judgmental Rationality**: Despite the fallibility of our knowledge, it is possible to make rational judgments about the validity of different claims. This involves critically evaluating evidence and considering the broader context in which it was produced (161).

Application in Guideline Committee Discussions

In the context of healthcare guideline development, adopting a critical realist perspective involves looking beyond the surface-level recommendations to understand the underlying mechanisms and contextual factors that the committee used to make their decisions and how they were influenced by their beliefs and preconceptions. This approach can provide deeper insights into how guideline committees discuss and interpret evidence to make informed recommendations.

Research design

This research aimed to explore how (and whether) guideline committees use qualitative evidence to formulate recommendations. A thematic analysis approach was used to systematically identify, analyse, and interpret patterns of meaning within qualitative data. Thematic analysis is a widely used qualitative research method that allows for the identification, analysis, and interpretation of patterns, themes, and meanings within data (162). This approach is particularly suitable for this research as it enables a detailed examination of how guideline committees engage with evidence to inform their decision-making processes.

Participants

Given the diversity of roles on NICE guideline committees, and the three broad areas in which NICE produces guidelines, a purposive approach to sampling was used. Purposive sampling better matches the sample to the aims and objectives of the research, thus improving the rigour of the study and trustworthiness of the data and results (163).

The study aimed to recruit a total of 12 participants from three committees (one clinical committee, one public health and one social care), which was deemed sufficient to provide a comprehensive view across the various types of guidelines and roles within the guideline committees. The sample included both members of the NICE guideline committees and NICE

technical staff who had supported these committees by conducting QES and presenting the results to the committees.

Specifically, the study invited six members from NICE guideline committees that had considered QES as part of the evidence base for their guidelines, and six members of NICE staff who had conducted QES and presented them to NICE committees within the past 18 months. To ensure a diverse range of perspectives, the recruitment aimed to include a variety of roles from the committee, including Chair, medical committee members, non-medical committee members, and lay (service user) members. This diversity was crucial for capturing a holistic understanding of the committee dynamics and decision-making processes and to better understand whether different professional groups had differing attitudes and beliefs about the value of qualitative evidence.

Recruitment Process

Committee members: Committee members were recruited through a self-selection process. The researcher used internal records and networks to identify three committees (one clinical committee, one public health and one social care) that had recently considered one or more QES as part of developing a guideline (in the previous 18 months) and negotiated access to these committees. A participant information sheet (appendix I) was distributed to members of the three committees as part of the paperwork that is circulated in advance of a scheduled committee meeting. During the committee meeting, in addition to the usual agenda, committee members were given a brief (<10-minute) PowerPoint presentation by the researcher to introduce the research study and provide an opportunity for committee members to ask questions. Committee members interested in participating were asked to follow up with the researcher by email.

Technical staff: Technical staff members who had conducted and presented QES to the committees in question were contacted via email and provided with a participant information sheet. They were invited to participate and given the opportunity to ask questions about the project before making a decision. This approach ensured that potential participants were well-informed and could make an informed decision about their involvement.

All participants were required to sign a consent form (appendix J) prior to the organisation of the interviews, ensuring ethical standards and participant understanding of the study's aims and procedures.

Data collection methods

Data for this study were collected through a series of semi-structured interviews, conducted online via Google Meet, and recorded using the inbuilt functionality. Participants were invited to partake in semi-structured interviews lasting between 45 and 60 minutes. The purpose of these interviews was to gather detailed data on the processes involved during the meetings where QES were presented and discussed, as well as the participants' perceptions of the QES and its influence on the recommendation-making process. Only the interviewer and participant were present during the online interviews.

This approach was chosen to facilitate in-depth exploration of participants' experiences and perceptions while allowing for flexibility in the conversation flow. In-person meetings would have been preferred but data collection started during the COVID-19 pandemic and University of Sheffield interim rules for researchers prohibited in-person interviewing. Data collection continued until 2022, but interviews remained online. Semi-structured interviews are particularly effective for qualitative research as they provide a balance between structured questions and the freedom to explore emerging topics in more detail (164).

Data analysis

The recordings of the interviews were transcribed verbatim by the author. Transcripts were uploaded to NVivo (118) and were coded using an emergent approach, with codes being allocated to units of meaning as they arose in the transcript rather than according to any pre-determined structure, however, initial decisions about coding were partly based on issues highlighted during the systematic review of the methodological literature and on the content analysis undertaken earlier in the PhD process. Once coding was completed, codes were read and re-read and organised into groups of meaning in the context of the research question. Critical discussion of the data and emerging themes between the researcher and supervisory team led to refinement of the emerging themes and the aggregation of those

themes into useful units of meaning. Groups were aggregated until further aggregation was not meaningful or diluted the detailed narrative of the lower-level groups.

Thematic analysis was conducted following the six-step process outlined by Braun and Clarke (165)

1. **Familiarisation with the Data**: Immersing oneself in the data by reading and re-reading the transcripts and documents. Initial notes were taken to capture preliminary ideas and patterns.

2. **Generating Initial Codes**: The data were systematically coded to identify significant features. Coding was done manually and with the aid of qualitative data analysis software (NVivo (118)) to ensure thoroughness and consistency.

3. **Searching for Themes**: Codes were collated into potential themes. This involved organising codes into broader patterns that captured the essence of the data. Themes were identified based on their relevance to the research questions.

4. **Reviewing Themes**: Themes were reviewed and refined to ensure they accurately represented the data. This step involved checking if the themes worked in relation to the coded extracts and the entire data set. Themes were merged, split, or discarded as necessary.

5. **Defining and Naming Themes**: Each theme was clearly defined and named. This involved writing detailed descriptions of each theme, including what was unique and interesting about them. The aim was to ensure that each theme told a coherent story about the data.

6. **Writing Up**: The final step involved weaving together the themes into a coherent narrative. This included a detailed analysis of each theme, supported by relevant data extracts. The write-up also discussed the implications of the findings for guideline committees and their use of evidence.

Furthermore, in line with the principles of thematic analysis as set out by Braun and Clarke, this study did not aim to calculate how many participants or items of data are necessary to achieve data saturation or ‘information redundancy’.

They argue that

> *although the concepts of data-, thematic- or code-saturation, and even meaning-saturation, … are not consistent with the values and assumptions of reflexive TA*

and therefore

> *recognise that meaning is generated through interpretation of, not excavated from, data, and therefore judgements about 'how many' data items, and when to stop data collection, are inescapably situated and subjective, and cannot be determined (wholly) in advance of analysis.* (134).

Ethics

Ethical approval for the study was obtained from the University of Sheffield's Ethics Committee (Reference number 035876 – see appendix K). All participants signed a consent form before the interviews commenced, ensuring both their voluntary participation and understanding of the study's aims and procedures, and their right to withdraw at any time. The confidentiality and anonymity of participants was maintained throughout the research process, in line with the University's guidelines and the General Data Protection Regulation (GDPR), including the anonymising of specific material in transcripts that could identify a participant (for example, reference to a specific guideline or workplace).

All data were stored on a secure University drive and managed in line with a Data Management Plan approved by the University ethics committee as part of the ethical submission. The Data Management Plan ensures that adequate care is taken with all data. It considers what data is collected, what documentation or meta data will be generated, ethical and legal compliance, storage, and back-up of data, keeping data, sharing data, and ensuring a named person is responsible for data management. Once transcribed and checked, original recordings were deleted so that only anonymised transcripts were retained.

Conducting Interviews on Google Meet

In accordance with the University of Sheffield's regulations for conducting online interviews, all interviews were conducted via Google Meet with the audio on but with the camera

turned off. This platform was selected for its accessibility, ease of use, and compliance with data protection standards as recommended by the University. Participants were provided with detailed instructions on how to join the online interview, ensuring a smooth and efficient process. The process of conducting interviews without video made the interview process more challenging at times since the interviewer and participants were unable to respond to visual cues that are part of normal interaction. However, it is to be assumed that the decision was made to preserve anonymity and to make it broadly equivalent (in terms of record keeping) with a face-to-face interview, which would normally have an audio record only.

Prior to the interviews, participants received an additional copy of the participant information sheet outlining the study's aims, procedures, and ethical considerations and a copy of the topic guide so that they could think about their responses in advance of the meeting. They were also given the opportunity to ask any further questions and provide informed consent. The interviews were scheduled at mutually convenient times, and each session lasted between 45 and 60 minutes.

Topic guide

In a systematic methodological review, five key steps were identified to develop robust topic guides: (1) identifying the prerequisites for using semi-structured interviews; (2) retrieving and using previous knowledge; (3) formulating the preliminary semi-structured interview guide; (4) pilot testing the guide; and (5) presenting the complete semi-structured interview guide (164).

Semi-structured interviews are appropriate for studying people's perceptions or opinions (166), in situations where the researcher has some previous knowledge of the phenomenon (164) and therefore were considered an appropriate approach to data collection (step 1). The format allowed the researcher, and supervisory team to contextualise the findings of the previous studies in Chapters 2 and 3, and to apply their own experience and expertise in designing a prototype interview schedule that was intended to encourage interviewees to talk about their opinions, perceptions, and experiences of using QES in guidelines (step 2 and 3). The prototype topic guide covered seven areas:

1. **Participant background in qualitative research.** This was felt to be important because participants who are more familiar with qualitative approaches to research may have different perspectives on its value than those who are more familiar with quantitative research. A key paper in the content analysis chapter (109) highlighted the importance of training to be able to understand and undertake QES.

2. **Participants experiences of hearing about the QES**. An opportunity to discuss the presentation of the QES, both as a written evidence review, and as a presentation by technical staff to the committee, as well as asking further about the general reaction to the QES - how the participant reacted to it as an individual, or as part of a committee.

3. **The balance between qualitative and quantitative evidence.** The purpose of this topic was to discuss the QES in the broader context of the evidence for an intervention/service as a whole. To discuss whether there was also quantitative or other evidence and to explore the roles played by the different kinds of evidence.

4. **The discussion of the evidence.** To elicit participants views and opinions about how the committee discussed the evidence and how they managed the interface between qualitative and quantitative evidence, especially where there was inconsistency.

5. **Recommendation-making.** Discussion about whether the committee made any recommendations on the basis of the evidence they considered. If so, was it the quantitative or qualitative evidence that drove the recommendation, and what was the value given to each type of evidence as part of the recommendation-making process.

6. **Possible improvements.** Given the participants experiences as discussed above, did they have ideas or suggestions about improving the process of considering QES as part of the evidence base for guideline recommendations. What would have made their job as committee members or as technical staff easier?

7. **Anything else?** All participants were offered the opportunity to address any issues they had expected to come up during the interview, or that they felt were important but had not had the opportunity to introduce during the interview.

The topic guide was populated with some prompts for interviewees and the interviewer and was piloted on the first three interviews to allow for any changes that might be needed to encourage the flow of the discussion and to ensure consistency across interviews. The aim was to allow for the exploration of individual experiences and insights and helping to maintain focus on the research objectives while providing the flexibility to probe deeper into relevant areas as they arise during the conversation (step 4) (167). Following piloting, the topics were not changed, but the prompts were updated to clarify the intent of the topics.

The topic guide was circulated to participants in advance of the interview to give them an opportunity to reflect on the questions before the interview. Sharing interview questions in advance with participants in qualitative research interviews offers several advantages. It can enhance the quality and accuracy of responses, as participants have time to reflect on their experiences and provide more thoughtful answers (step 5).

Data recording and transcription

Interviews were recorded with the participants consent using Google Meet's recording feature (which informs all participants when recording is in progress). The recordings were securely stored on a private, password-secured University of Sheffield drive in accordance with the University's information governance requirements and the Data Management Plan and later transcribed verbatim for analysis. Transcription was conducted by the researcher, ensuring accuracy and familiarity with the data. The transcripts were anonymised to protect participants identities. Brief field notes were made by the interviewer as reminders to follow up interesting lines of conversation. These were treated as confidential waste and destroyed on completion of the interview.

Ensuring trustworthiness and rigour

Trustworthiness and rigour in qualitative research are essential components in ensuring the credibility and reliability of the findings. Authors such as Lincoln and Guba have identified a number of key elements to consider (9):

1. **Trustworthiness**: This concept is often broken down into four criteria: credibility, transferability, dependability, and confirmability.
    - **Credibility**: This refers to the confidence in the truth of the findings. Techniques to enhance credibility include prolonged engagement, persistent observation, triangulation, and member checking (9).
    - **Transferability**: This involves the extent to which the findings can be applied to other contexts. Providing thick descriptions of the research context and participants helps others determine the applicability of the findings to their own settings (168).
    - **Dependability**: This criterion involves ensuring that the research process is logical, traceable, and documented. An audit trail and peer debriefing are common methods to enhance dependability (169).
    - **Confirmability**: This ensures that the findings are shaped by the participants and not researcher bias. Reflexivity and external audits are techniques used to achieve confirmability (170).
2. **Rigour**: Rigour in qualitative research refers to the strictness and precision with which the research is conducted. It involves maintaining a systematic approach to data collection, analysis, and interpretation.
    - **Systematic Approach**: Ensuring a systematic approach involves clear documentation of the research process, including data collection and analysis methods. This helps in maintaining consistency and transparency (171).
    - **Triangulation**: Using multiple data sources, methods, or investigators to cross-check data and interpretations enhances the rigour of the study.

- **Reflexivity**: Reflecting on one's own biases and how they may affect the research process and findings is crucial for maintaining rigour.

Conclusion

This chapter sets out the methodological approach taken to the primary qualitative study whose results are described in the next chapter. It sets the framework within which the research was conducted and by doing so seeks to ensure that the results are trustworthy and rigorous, and that the data collection and analysis were done in a way that protected the anonymity and privacy of the people who were interviewed. The following chapter describes the results of the interview and analytic process.

CHAPTER 5: RESULTS OF THE PRIMARY QUALITATIVE STUDY OF COMMITTEE MEMBER AND TECHNICAL STAFF VIEWS AND PERCEPTIONS ABOUT QES IN GUIDELINE DEVELOPMENT

Introduction

This chapter presents the results of the primary qualitative study described in the previous chapter. The purpose of the study was to better understand the views and perceptions of technical staff, committee experts and committee lay-members regarding how a QES contributes to committee discussions, and to the process of making recommendations.

Participants

A sample of 12 participants were recruited for the qualitative study. The participants were members of three NICE guideline committees who had considered qualitative evidence syntheses as part of their guideline development process, and NICE technical staff who had produced the syntheses and supported the committees in making recommendations.

The committees were selected pragmatically and for maximum variation. NICE guidelines cover clinical medicine, public health, and social care. One recent committee for each of these three areas was selected from which to sample participants. It was anticipated that attitudes towards the use of qualitative evidence might vary by field, with public health and social care professionals being more familiar with qualitative evidence than clinical professionals.

Within each committee, all members were eligible to take part. A short (<10-minute) presentation about the study was given by the researcher at the start of a target meeting, with the opportunity to ask questions. This was followed up by an email containing the information sheet and invitation to take part. Committee members who responded to the email were invited for interview. The intention was to purposively select participants to achieve diversity in committee role (Chair, member, lay-member) and profession (medical, non-medical, lay). However, a low volume of responders meant that all those who replied to the invitation were interviewed. Technical staff were recruited similarly via personal email invitation if they had participated in the qualitative synthesis presented to one of the three

selected committees. See Table 6 for brief details of interviewees. No interviewees refused to participate or dropped out following recruitment.

Each participant attended a single interview lasting between 45 and 60 minutes. In an effort to minimise the impact on busy participants, follow up interviews and participant validation (by commenting on or checking transcripts) were not undertaken.

Table 6 Overview of participant characteristics

Committee	Individual*	Committee role	Professional role
Clinical	CM011	Committee member	Non-medical
Clinical	CM012	Chair	Medical
Clinical	CM013	Lay-member	Public
Clinical	TS014	Technical (senior)	NICE staff
Pub Health	CM021	Committee member	Medical
Pub Health	CM022	Lay-member	Public
Pub Health	TS023	Technical	NICE staff
Pub Health	TS024	Technical	NICE staff
Social care	CM031	Committee member	Non-medical
Social care	CM032	Committee member	Medical
Social care	TS033	Technical	NICE staff
Social care	TS034	Technical (senior)	NICE staff

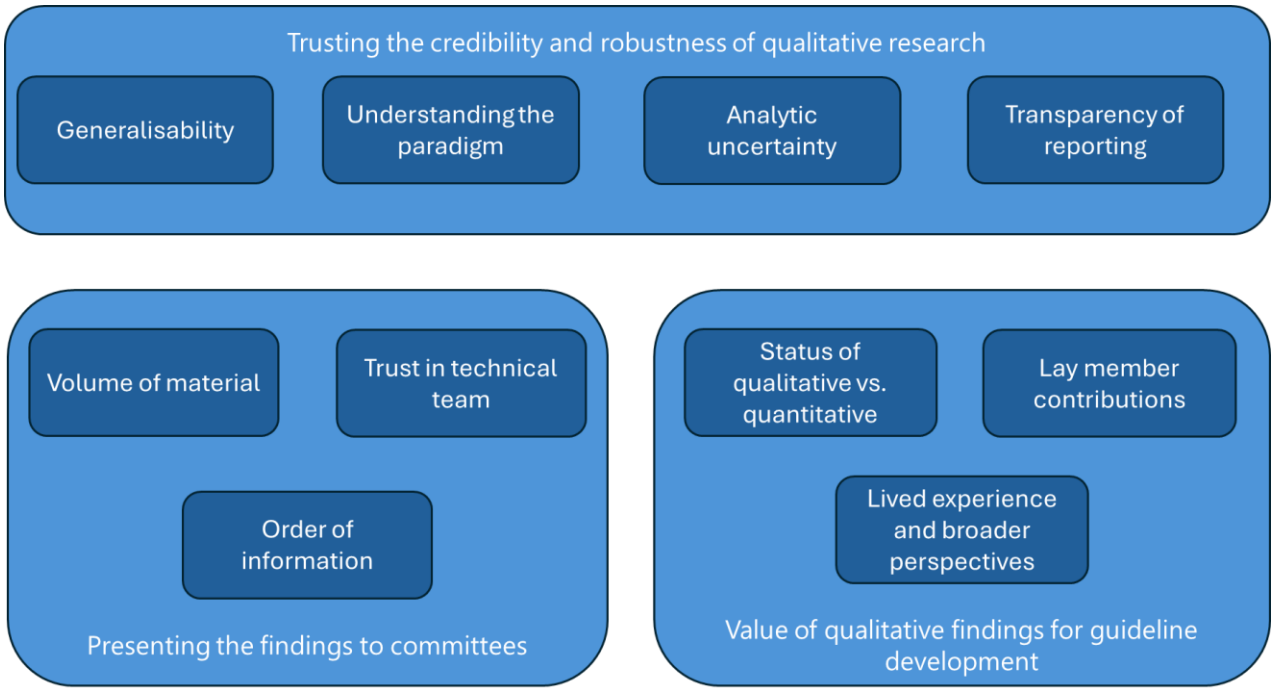
* The prefix CM denotes committee members; TS is technical staff.

Themes

Text from the interviews was analysed in-depth to identify recurring ideas or views amongst participants. Analysis of data from the interviews indicated a number of recurring themes. Three principal elements appeared to influence the way that QES are used/inform guideline development. Firstly, trust in the credibility and robustness of qualitative research and participant understanding of it. Secondly, the way that the findings of a QES are presented to a committee. Thirdly, perceptions regarding the value of QES during guideline development. A number of factors or sub-themes appeared to underpin each of these principal elements. Although the themes are presented separately, there is a fundamental

interplay between both the overarching themes and the sub-themes. This is particularly notable with respect to theme 1 (the credibility and robustness of qualitative research), where all of the sub-themes of theme 2 (presenting findings to the committee) and part of theme 3 (the value of qualitative findings for guideline development) can be traced back to the fundamental lack of familiarity and lack of confidence in qualitative research. A diagrammatic representation of the themes and sub-themes showing the structure of the analysis is presented in Figure 9.

Figure 9 Themes and sub-themes



Interlinkages between the themes will be explored in the next chapter. This is followed by an in-depth discussion of the themes and sub-themes. Throughout the text, inset italics are verbatim text extracts from participants. These are anonymised but the person's role and the type of committee are indicated. The text extracts have been selected to provide example data. Selection of verbatim text extracts for each theme was based on ensuring representation of individuals in different roles, with attention paid during analysis to any differing views or perceptions.

Trusting the credibility and robustness of qualitative research

Introduction to the Theme

Both committee members and technical staff expressed concerns about the credibility and robustness of QES findings. In the data, trust and credibility seemed to be linked to a lack of familiarity with the qualitative paradigm ("Understanding the paradigm"), this in turn was linked to a sense of unease about the analyses in the QES ("Analytic uncertainty") and meant that committee members felt that they would value clearer, more detailed reporting ("Transparency of reporting"). There were also attempts to apply positivistic notions of reliability such as sample size and representativeness ("Generalisability"). Each of these will now be explored in turn.

Understanding the Paradigm

In spite of the work that has been undertaken to develop methods for QES over the last decade (for example by the Cochrane Qualitative and Implementation Methods Group (138) and by the GRADE-CERQual working group (110)), committee members expressed uncertainty about the robustness of QES when used to underpin guideline production.

The uncertainty seemed to be related to limited understanding of the methods and underpinning principles of qualitative research generally. Most participants extrapolated their understanding of quantitative research within a positivist paradigm to qualitative methods, and from that perspective found it lacking.

> *I think, feel on confident ground when it comes to reviewing the evidence reports on the quantitative synthesis in that, you know, there's a nice forest plot there or there's that the…the numbers that are being presented are, uh, I guess, more objective in a sense. Whereas my experience of the way that our qualitative reviews tend to be presented is that within the kind of tabular data, we get a single quote, for example, that might illustrate a point.* (Medical Chair, clinical committee)

Similarly, technical staff were not always familiar with QES and needed additional support to undertake them.

> *I remember the very beginning. He [another member of technical staff] was quite a … he was kind of looking at it, like the clinical review question. I was trying to get him to look at it to be a bit more relaxed and not, you know, be so bound by the quality, clinical way of doing things.* (Senior technical staff, social care committee)

For at least two of the four technical staff, they had not carried out QES before.

> *So, some theoretical knowledge from my own kind of research training, although not actively participated in qualitative synthesis, kind of done masters related modules and kind of connected in a little bit that way.* (Technical staff, social care committee)

> *Yeah, there was the first proper ones. I guess the first ones w[h]ere I was doing anything like thematic synthesis. So yeah, so I had to kind of like familiarise myself with the Joanna Briggs Institute and figure out the different kinds of synthesis methods. It was all it was all kind of from scratch.* (Technical staff, public health committee)

Overall, committee members tended to regard QES, either overtly or by implication, as a secondary or supplementary form of evidence. A member of NICE technical staff summarised this by reflecting that quantitative data were seen as providing the essential evidence, with the qualitative data being almost complementary.

> *I feel like committees see the quant results as the real meat of the review, and qualitative as just adding the colour and adding "Oh, that's interesting. That's, you know, a nice, interesting extra detail, but we make our recommendations on the numbers because that is the proper science".* (Technical staff, public health committee)

Some participants applied markers of methodological robustness from quantitative reviews with qualitative reviews. This was true of committee members but was also true of NICE staff who had undertaken QES for NICE.

> *But again, then bias creeps in. They feel like more like you're cherry picking the interesting bits, but I think that might be a way to get the most relevant and appropriate data.* (Technical staff, social care committee)

It is notable also that when participants spoke about specific QES that they had been presented with as part of NICE committees, they were typically 'barriers and facilitators' type reviews, which may have limited their understanding of the paradigm. One member of NICE technical staff acknowledged this as a possible mistake.

> *We did what we shouldn't have done in terms of how you phrase the questions, I think we just kind of did a lot of barriers and facilitators. We were still learning a lot at the time. So, we just kind of jumped in with qualitative sounding questions, which would be something along the lines of You know, what are the barriers and facilitators to….* (Technical analyst, social care committee)

A committee member (on a different committee) also commented on the inter-relatedness of barriers and facilitators in those kinds of QES.

> *And then what are the barriers and facilitators, for [topic of the guideline]? So, we've got the sort of the mirror pair...* (Medical committee member, public health committee)

Participants had very mixed views on the content of QES, and of qualitative evidence generally, beyond the concerns about generalisability mentioned above. For example, one lay member from a science background talked about their familiarity with quantitative, but also how they could see the importance of qualitative evidence.

> *Yeah, I really do think it's equally important to have qualitative as quantitative evidence with a NICE guideline, or that within guidelines. And you know, I say this is someone with a master's degree, my undergraduate project research within statistical physics. So, I have no vested interest, you know, academically, in qualitative, my backgrounds in quantitative, but from that, I know that there's so many limitations of just numbers, and it doesn't kind of give you the whole picture. I mean, if we're just numbers, we could do everything by computers, you know, we are humans, we need to think of things through different lenses.* (Lay-member, clinical committee)

On the other hand, one person (a medic from a public health committee) explained at length that they perceived qualitative research on patient perspectives to be a kind of 'idealism' because it reflected a reality that was implausible for NHS services. The examples

used were weight management services or mental health services where service users might reflect in interviews that they did not want their weight or their mental health to be the focus of the appointment because it made them feel stigmatised. The interviewee pointed out that the purpose of the appointment was to discuss those issues and that reviewing evidence that was not compatible with support service configurations was unhelpful.

Transparency of Reporting

Participants agreed that in some ways, QES were much harder to make sense of as a reader than quantitative systematic reviews. They were opaquer and wordier, and lacked simple tables and figures that they associated with quantitative systematic reviews.

They highlighted the importance of transparency in systematic reviews and in QES, noting that it is easier to trust a synthesis if the derivation of the content is clear and can be traced from themes back to individual studies, and they generally did not think this was possible from reading the QES reports that were sent out to them in advance of the meeting. This meant that the bulk of the explanatory work needed to be done during presentations of the evidence to the committee. This often led to complex and very detailed presentations of QES from technical staff.

> *they were presented in a very detailed way, so that we really... understood what the what the particular project was about.* (Lay-member, public health committee)

or

> *I mean, we do go through it in quite a detailed way. [Int: Sure.] But I think that's important to do, to do it in the way it's been done. I can't really think of how it might be presented any better.* (Non-medical committee member, social care committee)

Although participants recognised the complexity of the presentation, they felt that it was reassuring and helped them to understand the complexity of the synthesis (see also 'complex presentations' below).

Analytic Uncertainty

Participants reported that it was difficult to understand the analysis process in QES. It was often difficult to understand how the themes reported in the primary studies mapped onto the themes reported in the QES, and this was often not possible to track meaningfully through the QES reporting. Because of this perceived lack of transparency, they generally thought that more detail was necessary in a QES than in a quantitative systematic review. Their inability to get easily from QES themes to the themes in the original data meant that a step-by-step presentation would give them more confidence in the themes in the QES, as would more detail about the methods of synthesis used. In spite of wanting detail, some responders commented that the detail that was included in NICE QES was too fragmented, with the information spread across various tables and appendices.

> *I think there's you know, there'll be a summary table, for example, of a description of, you know, that a study with focus groups and questionnaires in a certain set of [settings] in London boroughs. [Int: Yeah], but that's separate in a sense to the table that presents the findings, which is separate again to the table, to the diagram that might try and pull that together in terms of its… its synthesis. And so, it takes work to… to assimilate all of that information. And …and I guess one of the, and I think probably takes more work than it does for the way that the quantitative evidence is presented. So, I think that's probably one of the challenges.* (Medical committee member, social care committee)

Committee members noted that they could quickly match results from quantitative systematic reviews with the original papers (as long as they had access to the journal), but that this was much more difficult with qualitative papers, not least because they are often longer and lack tables and figures that make checking easy.

> *…colleagues sat around the table will look at the study that's being presented and reappraise it for themselves. [Int: Yeah], in order to...to have great confidence and they can do that real time very quickly with the quantitative work. And I think it's nigh on impossible to do that real time and quickly with the qualitative work, because I need to… I need to read a 20-page paper. [Int: Yeah]. As opposed as opposed to skim read a six-page paper.* (medical Chair, clinical committee)

This uncertainty seemed for some committee members to be further confounded by the sheer volume of themes that were included in some QES, especially during in-committee presentations. Technical staff agreed that the volume of material and papers that they had to synthesise could be prohibitive and they had experienced not only how this could overwhelm committees but were unsure what a solution could be.

> *But whether we needed something that then rather than presenting all the same themes forever, you know, again, and again, and again, maybe we needed a higher-level integration that was like, these are the themes that have come up for everybody. And then these are the separate themes.* (Senior technical staff, clinical committee)

Other participants agreed that the lack of specificity of some QES findings means that there was no meaningful framework on which to make recommendations, thus limiting the use of them in guideline development. The topics covered in QES they had seen tended to be quite broad brush in their approach rather than related to as specific intervention or service.

Generalisability

A common limitation mentioned was the lack of generalisability (transferability in qualitative terms) from the small number of participants.

> *And of course, the limitation of qualitative data is always, you know, how representative it is* (Non-medical committee member, clinical committee)

The committee member comments in this area reflect broadly voiced concerns about the ability to extrapolate from primary qualitative research to broader populations given how embedded in context qualitative data are. They also reported concerns about issues like applicability to the UK.

> *Some members of the committee would similarly feel slightly uncertain as to how widely applicable the findings might be or the area that we're discussing might be. You know, how can we extrapolate from a, you know, it's relatively small and it comes down to numbers again, is a relatively small population of the, you know, a series of focus group interviews or a series of semi-structured interviews in one population, in one [site] to extrapolate that out … And therefore, if we take that at face value, then,*

> *yes, we could make a recommendation based on that. But digging below that level, actually, how widely applicable is [it]?* (Medical Chair, clinical committee)

Participants used this concern to justify a position that qualitative data should primarily be used alongside quantitative data and should be moderated by the opinion of the experts on the committee, which led several participants to propose that the results of a QES should only inform recommendations if the committee agreed that the findings were 'correct' (i.e., that the findings matched their personal experience). In general, it was felt that QES might only be useful as supporting evidence, and that committee consensus was needed to gauge the credibility of QES findings.

> *And of course, the limitation of qualitative data is always, you know, how representative it is. So, I think what's important is that it's a combination of quantitative, qualitative and expert opinion from the committee in this particular context that I think is most valuable* (Non-medical committee member, clinical committee)

This also seems to relate to participants not having a strong grounding in qualitative paradigms, and therefore applying their understanding of statistical generalisability without realising that while sample size is crucial for statistical generalisability in quantitative research, qualitative research prioritises depth, context, and the richness of data, making the concept of generalisability more about the applicability of findings to similar contexts rather than broad populations.

Presenting the findings to committees

Introduction to the Theme

A series of sub-themes relate to the ways in which committees are presented with the findings of the QES during meetings. In large part the issues highlighted seem to relate to two factors, firstly (as discussed above) the analytic uncertainty and unfamiliarity with the qualitative paradigm mean that the participants felt as if they needed further detail, which led to a high volume of material being presented in long and complex presentations, and

secondly the fairly broad approach often taken to QES at NICE. This led the participants to highlight that trust in the technical team was more important.

A smaller, but significant theme related to the ways committees had discussed qualitative evidence related to the order in which it was presented to the committee.

Volume of Material

Participants from both committee members and technical staff wrestled with ideas that related to the size of syntheses and the volume of themes. The analytic uncertainty and inability to easily track themes back to primary studies (discussed above) meant that committee members needed to hear a much greater level of detail to be confident in the synthesis.

Technical staff also recognised this and reported that it was difficult to summarise large QES (in terms of volume of themes and quotes) into manageable slide decks and presentations that gave the committee the right level of detail without overwhelming them with information.

> *it's very easy to just present a forest plot and go look at the plot. And everyone's used to looking at graphs of some kind, and it's all fairly, fairly straightforward to get them to interpret. But yeah, I've kind of had to hone the presenting for qualitative a bit. Because I know they just tuned out the first time I did it with [topic]. I had a lot of interesting stuff - It was a big review. And yeah, it was it was the last thing on the probably a Friday. It's one of those ones where everyone just kind of just wanted to give up and go home and I'm just like, like, I'm just, you know, that really annoying teacher who's still talking… And so yeah, I know they didn't engage very much with that at all (*Technical staff, public health committee)

and a senior technical staff member noted that it had been an issue with a junior member of technical staff.

> *I had to try and stop him from just putting them [themes and quotes] up and reading them word for word or, you know, just like try and summarise them a little bit, and then let them [the committee] read them. Because he did just want to read them straight out…it was just boring as hell.* (Senior technical staff, social care committee)

Technical staff also noted that they needed to be very familiar with the content of qualitative papers to be able to respond to committee questions. Quantitative content can usually easily be summarised but the contextual information about qualitative studies cannot.

> *Yeah, it's I think it involves quite a lot of background knowledge to be able to answer questions on it. Whereas quantitative I feel like everything you need is usually on the slide.* (Technical staff, public health committee)

Trust in Technical Team

The uncertainty that interviewees talked about with regards to the transparency of QES led several of them to discuss the impact of this on the technical team producing the reviews. They noted that because it was not so easy to check, they needed to trust the technical team more, and because of this the technical team needed to be better at presenting qualitative evidence to committees in a way that reassured them that it was robust.

> *the committee, I think, has to put their trust in the technical teams, possibly even more than they need to with the quantitative work.* (Medical committee Chair, clinical committee)

Though some participants said that they did trust the work of the technical staff.

> *Yeah, I've always found the technical analyst, really, really good, really clear. And particularly when they do the summary slides. And it's really helpful to see how the different the different studies relate to one another.* (Lay-member, clinical committee)

It seems from comments like this that participants were reflecting the uncertainty they felt about QES methods. Concerns are rarely raised about processes and methods for quantitative techniques such as meta-analysis or pairwise meta-analysis, and data presented by technical staff were largely accepted, even though some, or perhaps even many, committee members are as likely to be unfamiliar with the specific methods used to generate the statistics as they are with qualitative synthesis methods.

In addition to concerns that related to the content of QES, participants raised concerns about the 'black boxes' in the process of producing a QES, and about how much of that

needed to be based in trust in the skills of the technical team because of the volume of data and tacit knowledge they had of the included papers (see also 'analytic uncertainty' and 'volume of material' sub-themes).

Order of Information

One member of technical staff talked about the nuance of whether to present quantitative evidence first and then support and expand with qualitative evidence, or vice versa and whether this had an impact on the way that committees hear and use data. They reported one occasion when they had presented the qualitative evidence first.

> *… we led with the qualitative and made the thrust of the point that we were making, about what the barriers were, and what the problems were, and then treated the quantitative, as here are some attempts to address these barriers. Here's how they fit into what barriers they're trying to address. And here's what level of success they've had with that. And I think that changed the way committee sort of viewed it, because I got them thinking about the barriers first, and think and seeing the quantitative in terms of the qualitative rather than the other way around.* (Technical staff, public health committee)

This participant perceived that this ordering had made the committee think differently about the qualitative data.

Although no committee members directly picked up on this point, there are many instances where participants imply this is the case, for example they talk about qualitative data being used to 'back up' quantitative data, or being used to 'fill the gaps' in the quantitative review (see the next section for further discussion of the relative status of qualitative and quantitative evidence).

Another member of technical staff described an instance when a clinical guideline committee had presented the qualitative data first and where, in the end, the committee made recommendations based on qualitative data outlining barriers and facilitators to the uptake of the intervention under scrutiny. However, it was also noted that the presentation of the qualitative data had been a challenge, and the committee did not have a framework

to build the recommendations around because they lacked the detail of effective interventions from the quantitative data.

> *Yeah, we did make quite a lot of recs, based around the qualitative work. So, what we ended up doing was we did the barriers and facilitators and effective interventions for [topic]. We …, presented them to the committee made, we didn't really, we didn't actually make any recs at that point. Because the committee, we didn't have anything, you know, we didn't have a framework to put them in.* (Senior technical staff, clinical committee)

They had resolved the presentational issue with a mixed-methods approach to generate a kind of evidence matrix.

> *And then what we did was every time we had say a review on [topic]. We went through all the quantitative data. And then we pulled out all the qualitative themes from the different age groups that related to [topic] and brought them into the into that review …. And then we did a diagram, … where the evidence matched the key themes from the quant and the qual, … And we tried to bring them together. And then at that point, they made the recs based on the quantitative and also then took the qualitative into account.* (Senior technical staff, clinical committee)

It is notable in this description that the final statement refers to the primacy of the quantitative data, even though the participant response was framed initially around making recommendations based on QES (see also the section on credibility above and on value below).

Value of qualitative findings for guideline development

Introduction to the Theme

A final series of sub-themes relate to the value of qualitative evidence for informing committees' deliberations and recommendation-making. Participants almost universally agreed that QES were useful in improving representation, and that they could do this in two ways, firstly by focussing on the patient or service user voice, and secondly by giving some insight into the experiences of different groups (for example, people from different ethnic

or faith backgrounds, or people with disabilities) in a way that could make guideline recommendations more applicable to those groups.

Other than this, participants seemed only to have a vague idea of the value of qualitative evidence, but in general were very unclear what that value is in terms of making recommendations.

Status of Qualitative vs. Quantitative Research

It was notable that no participants recognised that there might be occasions where the qualitative evidence might be more pivotal to decision making than the quantitative evidence, even though in many cases they framed themselves as supportive of, or at least open-minded about qualitative evidence.

> *No, I think even what I would describe as the hard-core quantitative people [laughs] have, have found the qualitative evidence important as well. Yeah, I don't think I don't think there's been anybody that sort of just poo-poo'd it at all. It's all been... you know… I mean, that's... I mean, outside the committee I've certainly had those conversations with other people. But I haven't felt that at all because I felt that the committee can really kind of relate to what's coming out in the qualitative evidence.* (Non-medical committee member, clinical committee)

However, this same committee member goes on later in the interview to express surprise that NICE consider it sufficient for recommendation-making. The participant reflects that it might be seen as important if it is "all you've got". Views seemed particularly strong from the clinical committee.

> *I was quite surprised by how much, there..., how much store is put by it. Given that you're making recommendations. But, you know, in the absence of anything else, then it... it is obviously it's it becomes very important if that's all you've got for a particular area.* (Non-medical committee member, clinical committee)

Such views were not uncommon, even if sometimes they were conveyed subtly, for example implying that the core part of a guideline recommendation would only be built on quantitative data with qualitative data adding nuance and detail.

> *a feel for a lot of the recommendations of kind of, like the quantitative data was the [clothes] hanger and the qualitative data was the clothing draping over the hanger.* (Senior technical staff, clinical committee)

but even on committees where one might expect the committee membership to be familiar with qualitative data there was still a feeling that the qualitative data might have been useful only because there were little or no quantitative data.

> *There's a kind of a feeling that one of the reasons that qualitative evidence was useful, was because there was so little quantitative evidence. And I guess that the sequela of that, does that mean that if there were a lot more quantitative evidence, then the qualitative evidence would have been less or not useful?* (Technical staff, social care committee)

Around half of the participants did not conceptualise the possibility that evidence from a QES could be 'stand-alone' evidence and only ever referred to it as part of a mixed methods 'package' of qualitative and quantitative studies. This could partly be because their only exposure to QES had been as part of mixed-methods, or parallel qualitative and quantitative syntheses, because at the time of the interviews NICE was exploring mixed-methods synthesis in public health and social care.

Participants were of the view that generally, qualitative evidence alone would not normally be sufficient to make guideline recommendations:

> *So, I think making recommendations in isolation is rare. So, in a sense, I can't recall us being in a situation where we are using just one type of evidence review to make recommendations.* (Medical committee Chair, clinical committee)

Although this could be read as an implicit acceptance that qualitative evidence is valuable, in the majority of cases NICE recommendations are based on quantitative evidence alone, so this is unlikely to mean that recommendations would not be made on quantitative evidence alone.

Even though some participants believed those data could be useful for recommendation-making, it was mostly framed in terms of using the QES themes to support consensus

recommendations rather than the qualitative themes being the source or justification for the recommendation.

> *…rather than making a recommendation that came straight out of the qualitative evidence, they… they did use it to bolster what they believe to be true anyway and make a recommendation based on that.* (Senior technical staff, social care committee)

However, the sense was also expressed that the best recommendations grew from a combination of different kinds of evidence and from committee expertise and experience.

> *So, I think what's important is that it's a combination of quantitative, qualitative and expert opinion from the committee in this particular context that I think is most valuable…* (non-medical committee member, clinical committee)

The (mostly) tacit creation of a hierarchy of evidence by committee members was an area where, for several of them, the narrative of their interview broke down to some extent. Participants who claimed to value qualitative evidence were often valuing qualitative data only insofar as it supported their clinical expertise and experience. It was clear they would only make recommendations based on QES that matched their beliefs, whereas they were more likely to accept the quantitative data regardless of their previous experiences or beliefs in the area. This was not unnoticed by NICE technical staff (see quote from 'Senior technical staff, social care committee' above).

In contrast to this, technical staff agreed that having qualitative evidence that reflected committee consensus was helpful because the qualitative evidence could be used to add strength to what would otherwise be a weak consensus recommendation and could help with getting the recommendation approved by the guidelines quality assurance (QA) process.

> *I think it [QES] is adding, because I think the difficulty we've had in the past is like, you can't make recommendations on something you haven't found evidence on. So, I think now that we have, like, …it's not really showing us anything new or groundbreaking. But it's given us the opportunity to say this is a problem. It's been identified, and we're making recommendations. So, I'm hoping when we do get to the QA stage, there's less*

> *of that fight for that recommendation stay in. Whereas I feel like in the past, we really had to kind of fight our corner.* (Senior technical staff, social care committee)

Lived Experience and Broader Perspectives

Committee members and technical staff agreed that one of the main difficulties of writing guidelines lay in trying to understand the impact of potential recommendations on people from diverse groups, especially those where there are likely to be more barriers to care. They noted the potential of QES to begin to redress this balance because qualitative research often focuses on harder-to-reach people and uses sampling techniques designed to capture the widest range of voices possible. Often knowledge of these diverse groups was within the committee's experience, but they agreed that having it reflected in a QES gave weight to their clinical experience. This quote conveys a sense of things being different in different groups, and the value of the QES is that, rather than just 'state' something, they can cite evidence to support it.

> *I think with quantitative evidence, … it's very rarely that the equalities … groups are … covered. But there is a small possibility that they are covered in the qualitative evidence.... not that we found anything groundbreaking again for like, for example, BME [Black and minority ethnic] groups, but you know, there was a bit where … one of the papers, you know, said [that] different ethnicities, you know, consider weight differently. So culturally some, some cultures like people to be on the bigger end of the scale. So, like things like that. And again, it's a given, but it's nice to have it written down and like documented somewhere that we did find this and it was taken into consideration.* (Technical staff, public health committee)

Conversely, it was noted that this same variation could make it harder to understand how the evidence might relate to a general UK population, especially if the study were not a UK study.

> *…they might do a study in a sort of a poorer ethnic area, but the ethnicities they'd be interested in aren't necessarily the same as ours, and their societies not identical to ours. So, you do have a bit of is would this work here is, you know, it seems to work in general, but would that work with our populations? And that so that was a little hard*

> *to generalise when they didn't do it with the general population?* (Medical committee member, public health committee)

This concern about generalisability brings the analytic narrative full circle, to questions about the validity of QES based on its lack of generalisability, a concern that is not specific to NICE committee members, but also a discussion that has been taking place within qualitative research communities.

Several participants across all three committees took the view that in many cases the findings of QES reflected their lived experience as clinicians and healthcare practitioners, but this was formulated either positively (because it supported the validity of their personal experience), or negatively because it was perceived to be 'stating the obvious' (see below).

One interviewee (non-medical committee member, clinical committee) noted how the evidence 'resonated' with them, and could see it also resonated with other people on the committee. A member of the technical staff observed that this could be quite empowering for the committee.

> *And it kind of gave them the reassurance that we … you know, this isn't just us it's… it's… it's in the data as well that this is a this is a massive issue. So, I think that really helped them.* (Technical staff, public health committee)

But for some committee members this felt as though the QES was just repeating the obvious and telling them what they already knew.

> *Yeah, the trouble is, it often ended up being a theme, which was perhaps kind of understood anyway, in terms of like, you know, say, [the service user group] will be saying something like, we just want real relationships. We just want we don't want to feel like you're here because it's a job.* (Medical committee member, social care committee)

Lay-member Contributions

Non-lay-member participants took the view that the themes identified in QES could serve as a way for committee lay-members to validate their own lived experience.

> *And certainly, people who we've had on [committees] have undoubtedly felt reinforced in some of their views and their experiences by hearing, that sort of thing come from other people.* (Medical committee member, public health committee)

and believed that sense of validation could give them the confidence to speak confidently to the committee about their views.

> *We had a qualitative review on [topic]. And that was definitely where they [lay-members] came into their own for that particular meeting. I remember, all there were really contributing for that meeting…And yeah, they were they were really vocal at that meeting. I think that maybe the Chair was saying "no, this is this is your opportunity to contribute...". So, they all did step up and they all contributed.* (Technical staff, social care committee)

One Chair even suggested that if the correct lay-members were on the committee then QES was pointless because the lived experience of the (usually two) lay-members was sufficient to bring a qualitative perspective to the committee meetings.

> *people have the lived experience of what they're… what you're talking about. And you'd be pretty well…, well you'd be wrong, but also, you'd be pretty stupid not to listen to what they have to say, because that's very valuable. Whether the qualitative research synthesis goes above what those lived experience people will tell you anyway, I'm not convinced, personally.* (Medical Chair, clinical committee)

NICE committees most commonly contain two lay members who come from a background of lived experience, either because they live with a disease or condition, or because they are affected by it, for example they may be a carer or parent to someone with the condition. The unexamined assumption in the quote above that the lived experience of two people is as informative as synthesised data from a (relatively) much larger and diverse sample, across multiple studies and settings. This reflects a sense throughout this project that participants broadly equate QES and their own lived or clinical experience in terms of their status and value as evidence.

Sometimes the engagement of lay-members with the qualitative evidence was framed by committee members in a way that could be interpreted as saying that lay-members engaged

better with qualitative evidence because they did not understand or relate to quantitative evidence.

> *I think the qualitative evidence, my… my view from my committee, the qualitative evidence gave the lay members of voice because so far, we've looked at very much technical pieces of information where they can't really chip in there, you know…* (Medical committee member, social care committee)

Or

> *My guess is possibly the lay members would gravitate more towards the qualitative because they lived experience and we're here at lived experience. And my hunch is that the kind of scientists and medics might pay more attention to the quantitative, but I can't really think of an example.* (Lay-member, clinical committee)

Some lay members also seemed to imply that the QES offered them more opportunity to understand and respond to the evidence than the quantitative data, and importantly that it gave them insight into other people's lived experiences to consider alongside their own.

> *I guess one of the things I tried to see as a lay member is not just think about my own experiences but try to think of others. But again, it gives me a lot more to think about when it's taught me about other people's experiences, whether they resonate with me or not, compared to just more numerical values.* (Lay-member, clinical committee)

Conclusions

Having summarised the findings that emerged from the qualitative study, the next chapter digs more deeply into them, particularly looking at the interlinkages between them and their fit with existing literature about QES and guideline production, as well as with the findings from the earlier systematic review of the literature (Chapter 2) and content analysis (Chapter 3).

CHAPTER 6: DISCUSSION OF FINDINGS

Introduction

This chapter will provide further analysis and interpretation of the data and themes in the previous chapter and explore them in greater depth.

The chapter will then seek to build on that work to generate new understandings based on the three studies undertaken for this thesis. Specifically, this chapter will explore how the findings from the systematic review (Chapter 2) and the content analysis of NICE QES (Chapter 3) informed the design, interpretation, and contextualisation of the qualitative study (Chapter 4 & 5). This integration provides a richer understanding of the challenges and opportunities in using QES in guideline development.

The interdependency of the themes and sub-themes

The previous chapter developed a framework of three key themes, each with several sub-themes that was used as the structure for the presentation of the data from the primary qualitative study (Figure 10). However, reading the description of the themes will have made clear that although the structure has been presented in discrete sections, the interlinks and dependencies between each of the themes and sub-themes is much more complex.

Figure 10: Structure of presentation of themes

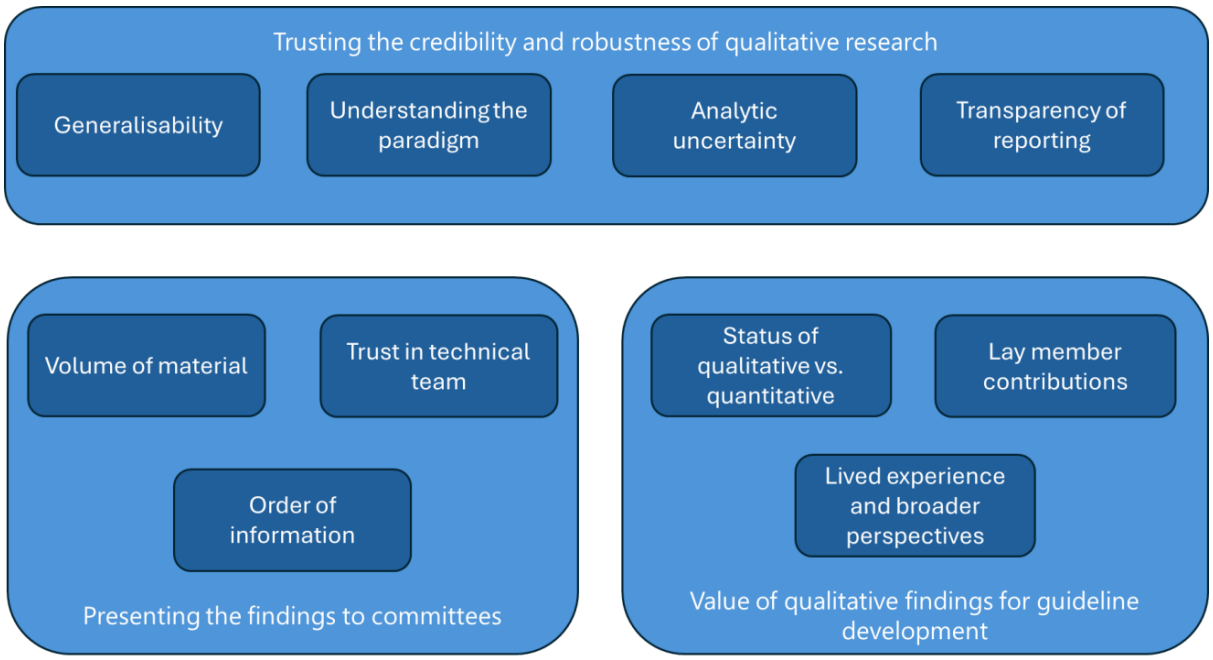
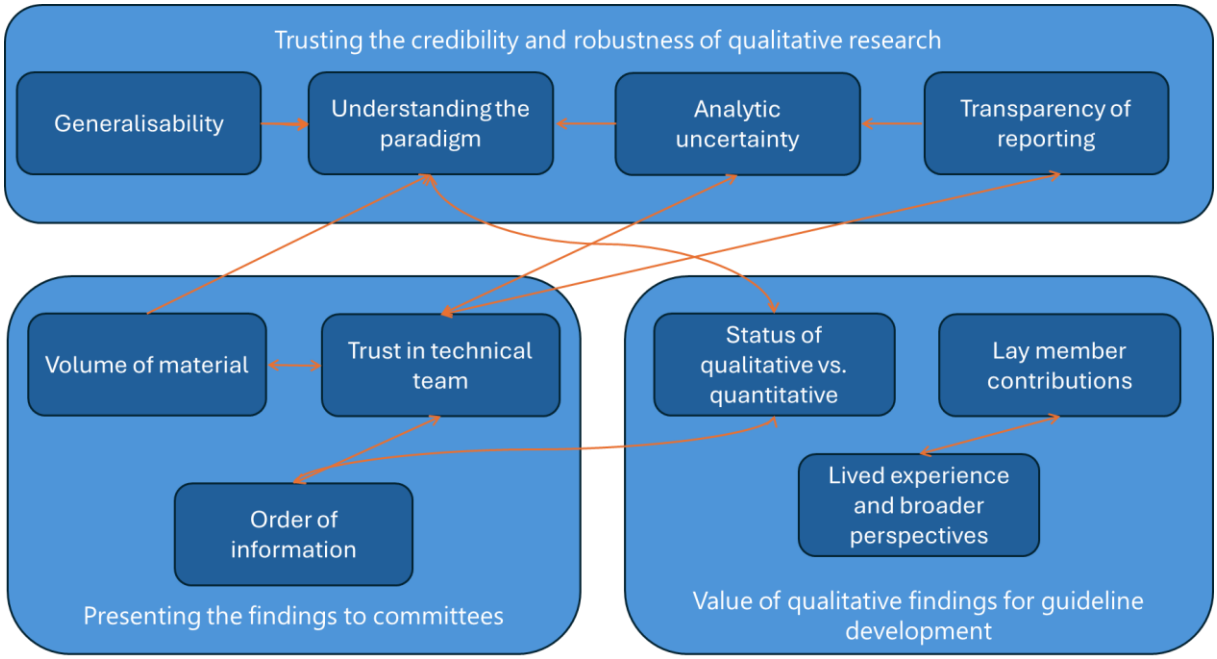


Figure 11 uses orange arrows superimposed on the thematic structure to demonstrate the key interdependencies between the sub-themes. This, of course, implies a high-level interdependency between the three overall themes (not shown with arrows for the sake of graphic simplicity).

The act of 'zooming out' seems to reveal that the key driver underlying many of the issues and sub-themes found by the analysis is the theme of 'trusting the credibility and robustness of qualitative research', and specifically in terms of the sub-theme about 'understanding the qualitative paradigm'. It seems reasonable to hypothesise that not understanding the qualitative paradigm could easily lead to 'analytic uncertainty' and concerns about 'generalisability'. These in turn would require more 'trust in the technical team', who might need to produce a larger 'volume of material' in an effort to produce transparent reporting to try to defuse the analytic uncertainty expressed by the committee members (and technical staff).

Figure 11: Key links between sub-themes



Interestingly, and a plausible mechanism of some of the interdependencies, is that broadly speaking, committee members seemed to be mostly unaware that they had limited understanding of the qualitative paradigm generally. This was not the case for technical staff, who commonly reported having no background and limited (or inadequate?) training in qualitative synthesis. It is also notable that very few NICE technical staff have any qualitative research experience, and it seems likely that experience in conducting qualitative research would be a useful skill for those planning to undertake qualitative evidence synthesis.

Central themes and their challenges

Trust and Transparency

Although they are themes in their own right, trust in the technical team and the transparency and robustness of the QES process are recurring issues that also seemed to underpin most themes in the analysis, and furthermore reflect issues raised by methodologists in the section on methods of synthesis in Chapter 2 about the need for transparency and to be able to see how themes were derived (77, 172, 173). The need for detailed, transparent reporting and the reliance on the technical team's expertise underscore the broader issue of trust in, and analytical uncertainty around, the qualitative evidence synthesis process discussed above. This trust is crucial for the committee to have

the confidence to produce recommendations, highlighting the importance of clear communication and robust methodological practices (137).

Committee members expressed concerns about the opacity of QES reports, which were often seen as more complex and less accessible than quantitative systematic reviews. This lack of transparency can undermine confidence in the findings and the recommendations based on them. Therefore, improving the clarity and traceability of QES reports is essential for building trust among committee members. There are challenges inherent in this:

1. QES are usually more 'wordy' than systematic reviews. Even with good use of visual tools (174), the use of quotes and of narrative sections in (for example) CERQual tables.

2. QES are often written by analysts whose skill is in quantitative systematic reviewing, and they may not be skilled in distilling qualitative information into succinct reports.

3. Committees' unfamiliarity with QES means that a more complete write-up may be required that includes greater detail than would normally be the case, despite knowing that committees are concerned about the complexity of the work and the volume of information.

These concerns echo the findings from Chapter 3, which showed that many NICE QES lacked detailed methodological reporting, particularly in areas such as synthesis methodology and theme derivation. This lack of transparency likely contributes to the analytic uncertainty and mistrust expressed by participants in Chapter 5. Furthermore, Chapter 2 highlighted the importance of reflexivity and transparent synthesis processes—principles that were often not evident in the NICE QES reviewed in Chapter 3, and which may partially explain the discomfort and scepticism voiced by committee members

Challenges of Presenting Large Volumes of Complex Data

Managing the complexity and volume of material in QES has been an issue not only for NICE, but also for other guideline producers such as WHO. Methods have been proposed to manage large numbers of inclusions in robust ways, and these could be easily adopted by NICE. See, for example, the chapter on selecting studies and sampling in the forthcoming Cochrane-Campbell Handbook for Qualitative Evidence Synthesis (175) and Ames et al's

(133) method for purposive sampling of potential includes for QES, where they sampled from their potential includes using a framework that prioritised studies undertaken in low- and middle-income countries, studies that scored highly for data richness, and studies that closely matched the objectives of their QES (a tool to assess data richness and thickness specifically in the context of QES has been developed (85)). Adoption of an approach like this might have prevented some QES of over 50 papers that were documented in the content analysis (Chapter 3) – although the content analysis also calculates a mode of four includes in NICE QES sampled there. The possibility of having concise or succinct themes that the committee see as relevant to the recommendations that they are discussing is further diminished by the inexperience of NICE staff in undertaking QES, with a tendency to use a vague or overly inclusive research question, which often leads to a vague 'answer' in the form of a large number of themes, many of which are not obviously relevant to the review question or decision problem. This issue was anticipated in the methodological literature reviewed in Chapter 2, which emphasised the need for focused review questions and purposive sampling to manage data volume. The content analysis in Chapter 3 confirmed that many NICE QES used broad or generic questions (e.g., barriers and facilitators), which may have contributed to the thematic overload described by participants in Chapter 5. Thus, the design and scope of QES (Study 2) directly shaped the experiences and perceptions of committee members (Study 3).

The difficulties in presenting and interpreting large volumes of complex qualitative data highlight a broader challenge in evidence communication (176). Effective presentation techniques and innovative methods for synthesising and displaying qualitative data might make it accessible and actionable for committee members.

Additionally, mixed-methods approaches to integrating qualitative and quantitative data in a coherent framework can provide a comprehensive view of the evidence (177) as well as formalising a more equitable weighting of the qualitative data in relation to the quantitative data. They can also provide a comparatively simple way to fit those data together in a way that can inform overall recommendation-making – a single way with which committees can then all become familiar (despite any limitations it might have), thus addressing many of the core themes identified in Chapter 5, such as analytic uncertainty, volumes of data etc.

The evidence from Chapter 2 showed that the ENTREQ reporting standard is commonly cited as the preferred standard for QES in guideline manuals and related papers although many QES use an adapted version of the PRISMA guideline or an adapted Cochrane review template. The Cochrane Effective Practice and Organisation of Care (EPOC) group published a template for Cochrane QES which is still available via an archive site even though Cochrane EPOC no longer exists[3]. This is also now available as a template in the RevMan software (178). The content analyses (Chapter 3) considered how well the NICE QES sampled match the ENTREQ reporting criteria, and it seems obvious that adherence to well established, robust reporting criteria will make a QES readable and transparent and ensure that the key information is included. Figure 5**: Number of QES meeting each ENTREQ reporting criterion** (in Chapter 3) demonstrates that there is good consistency within the ENTREQ criteria as to whether it is done well or poorly in NICE QES, with most criteria either reported by almost all, or alternatively by around half of the QES sampled. It was noted that the criteria in the framework where the level of QES reporting is very high are those criteria that duplicate steps in quantitative systematic reviews and are therefore familiar to staff who are predominantly quantitative systematic reviewers, as is the case with NICE technical staff. The ENTREQ tool itself is commonly cited as the key reporting standard for QES (Chapter 2) alongside the Cochrane review format and bespoke variations of the PRISMA standards for quantitative systematic reviews. A team from the universities of Stirling, Bangor and Sheffield are currently developing a PRISMA reporting guideline for QES that is due for completion in 2027 (67). They note that the ENTREQ criteria, published in 2012 pre-date many important changes in QES methodology, including the creation of GRADE-CERQual, furthermore, the development process for ENTREQ would not be considered robust by current standards. Other, recent reporting standards such as eMERGe are specific to particular forms of synthesis.

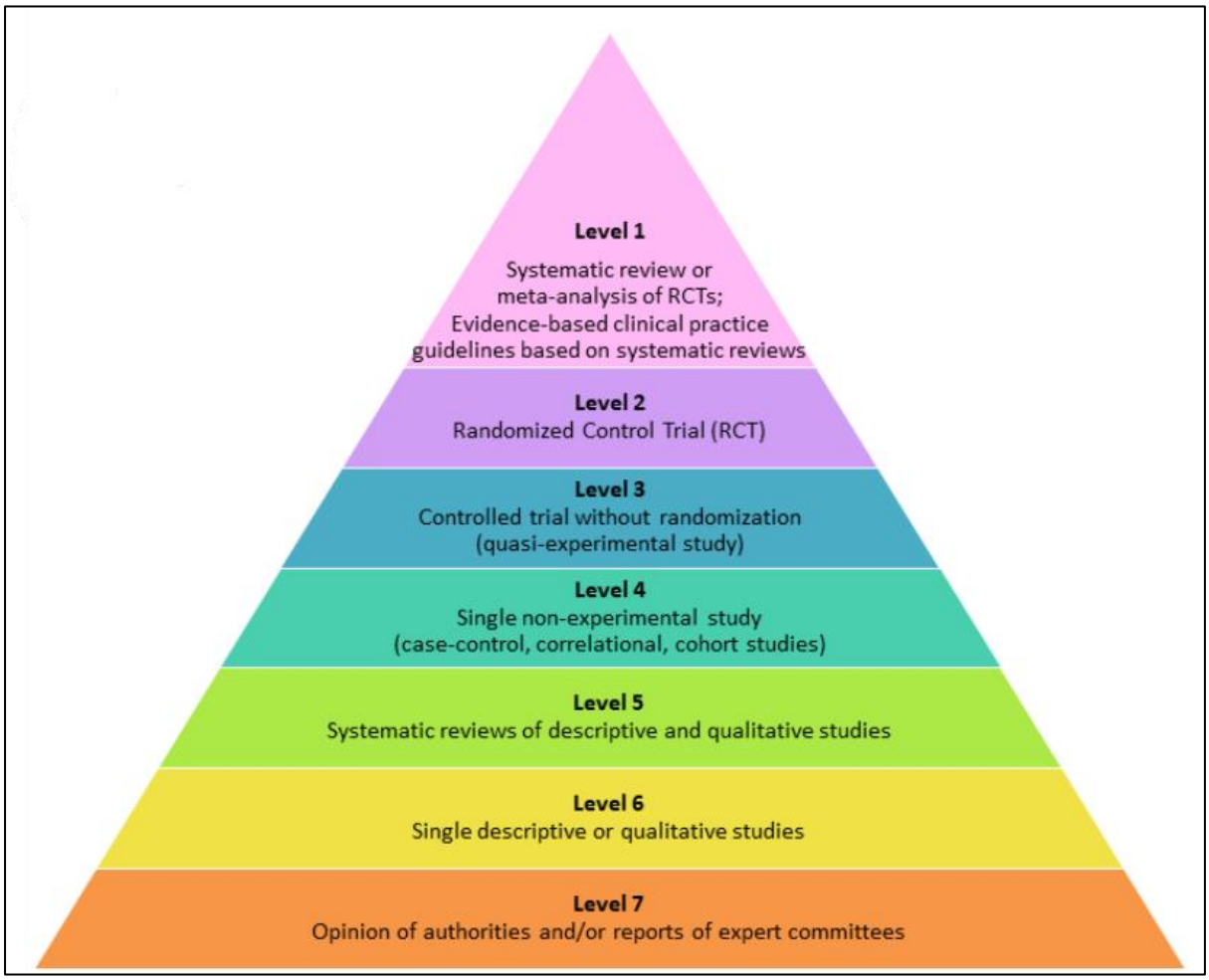
Status of Qualitative vs. Quantitative Research

Overall, guideline committee members and NICE technical staff tended to be critical of QES, either overtly or by implication. It was seen as inferior to quantitative evidence for being the basis of guideline recommendations, and by implication was often considered to be on a footing with, or even less robust than, committee consensus and the lived experience of lay

[3] https://zenodo.org/records/10050961, accessed 23/01/2025

members. This view matches neatly with the historical hierarchies of evidence that have been expounded in evidence-based healthcare (EBHC) for many years and can be found in almost any medical, nursing or allied professional textbook that discusses EBHC or its predecessor evidence-based medicine (EBM) – see Chapter 1 for a brief overview of EBHC. Figure 12 shows an example from a nursing textbook which places 'systematic reviews of qualitative studies' and qualitative studies just above expert opinion (179).

Figure 12: Hierarchy of evidence



LoBiondo-Wood G, Haber J. Nursing research: Methods and critical appraisal for evidence-based practice. St. Louis: Mosby Elsevier; 2014.

This perception aligns with the findings from Chapter 3, where QES were often used to address narrow or supplementary questions, reinforcing their secondary status. Chapter 2, however, presented a broader vision for QES, including their potential to inform scope,

identify values, and shape recommendations. The disconnect between this potential and actual practice may help to explain the ambivalence and undervaluing of QES observed in Chapter 5.

Interplay between quantitative and qualitative evidence

The integration of qualitative and quantitative evidence in guideline development is a complex process that requires careful consideration of the strengths and limitations of each type of evidence. In terms of its health and social care guidance, NICE (and most guideline developing organisations) have a very firm focus on 'what works?' in terms of the efficacy and effectiveness of drugs, surgery, or any other kind of intervention. In that context, it is easy to argue that historical hierarchies of evidence hold true, and that the best way to answer those questions is through a meta-analysis of randomised controlled trials. However, guideline recommendations are usually based on more than the effectiveness evidence alone. In a changing NHS (in the UK) context that has greater focus on shared decision making, patient preference and personalised care, the most efficacious intervention (in a clinical sense) is not necessarily the 'right' intervention for a particular patient. Currently, that contextualisation of the effectiveness evidence is primarily done by committee consensus. In many cases that might be entirely appropriate, and the resource implications of conducting QES alongside most or all effectiveness questions would be prohibitive, but equally, it seems remiss to undertake a 2 year guideline development process looking at (for example) diabetes care for people with type 2 diabetes, without starting that process by trying to ascertain how and why people with type 2 diabetes decide to engage (or not) with that care. Chapter 2 advocated for the use of QES to inform not just implementation but also the framing of review questions and the interpretation of effectiveness evidence. However, Chapter 3 showed that NICE rarely uses QES in this way, and Chapter 5 revealed that committee members often default to consensus or personal experience in the absence of clear qualitative evidence. This suggests a missed opportunity for QES to play a more central role in shaping guideline development.

A deeper analysis reveals a more nuanced interplay between qualitative and quantitative evidence. Qualitative evidence provides context and depth to quantitative findings, offering insights into the experiences and perspectives of individuals that numbers alone cannot

capture (180). This complementary relationship suggests that both types of evidence are essential for comprehensive guideline development. While the participants recognised this verbally, further enquiry often revealed this view was not deeply held and carried a sense of "talking the talk", in the sense of saying what they thought they should say, rather than what they actually believed, supporting the view that the challenge lies in ensuring that qualitative evidence is not merely seen as supplementary but as an integral part of the evidence base that seeks to address different questions (181).

The discussions about the credibility of QES focussed on the perceived limitations of qualitative research methodologies, their perceived inferiority to quantitative methods for providing useful evidence for recommendation-making, and about the perceived difficulties involved in understanding and unpicking that evidence. Committee members and lay-members also commonly talked about how the QES often matched their lived experience of working in, or of using, a service. This served both to reassure them that the QES reflected 'real-life' (and thus had value) but also made them question whether this was because the findings were so obvious that the QES was pointless.

In spite of its perceived limitations, participants were supportive of the use of QES during guideline development, primarily because it gave them a broader perspective on the evidence and could give useful evidence in areas where quantitative evidence would be less useful or was less plentiful. They were particularly convinced by the value of QES, both as a means to increase the involvement of lay members in the committee, and also as a means of shedding light on inequalities that might not be uncovered by the quantitative reviews.

Implicit biases and hierarchies

There seem to be implicit biases and hierarchies in how different types of evidence are valued by participants. Despite them stating support for qualitative evidence, they often portray it as secondary to quantitative data. As noted in the previous chapter, there was in some cases quite a dissonance between the ways that they consciously formulated their opinions about QES, and the contradictory unspoken assumptions that underpinned other comments, such as the clinical committee member previously discussed who spoke about the importance of qualitative evidence, but then later reflected that it might be seen as important if it's "all you've got" (Chapter 5 – status of qualitative vs. quantitative evidence).

This bias can influence decision making processes and the development of guidelines, suggesting a need for cultural change within committees to truly value and integrate diverse types of evidence (182).

The other, unexpected, implicit bias was found in the committees' comments and thoughts about the impact of the order of presentation of evidence for review questions where there was both quantitative data and a QES, or where there was an attempt to undertake formal mixed-methods integration between the quantitative and qualitative reviews. In spite of speculation about the importance of the order in which information was given, it was almost always formulated by committee members that the framework of a recommendation, and its raison d'être, was drawn from the quantitative evidence, and the qualitative evidence was used to flesh out or nuance the recommendation. Making recommendations only based on qualitative evidence was rare at best. No interviewees noted making any recommendations that were primarily based on qualitative evidence. This might in part be because the driver of review questions is the effectiveness element (as discussed above) but could also be due to the ways that NICE use or choose to undertake QES. The content analysis study presented in Chapter 3 notes that the largest proportion of NICE QES conducted during the 5-year period analysed for that study were reviews of 'barriers and facilitators', which do not readily lend themselves to stand-alone recommendations, but to adding context to a recommendation based on effectiveness evidence – "deliver [effective intervention] in a [facilitators] way". Reviews about the information needs of clients with a particular condition or diagnosis, or reviews of the care and support needs of people with a specific condition might plausibly lead to more direct recommendation-making.

It appears that if NICE want to produce more holistic guidelines, then work needs to be done to address the clear issues with QES as part of the evidence base. Although their processes remain far from ideal, it appears that other organisations, for example the WHO, who aim to produce more holistic guidelines, may have achieved greater evidence equality, perhaps primarily through their focus on the WHO guideline panels' requirement to accept (or not) each individual finding, both qualitative and quantitative (personal communication with Chris Carroll, PhD [December 2024]).

Value of Lived Experience

The integration of lived experience through QES and lay-member contributions is highly valued by the participants in the qualitative study, yet there is tension between this and the perceived need for generalisability. This suggests a need to better articulate and validate the role of lived experience in evidence synthesis, ensuring it is seen as complementary rather than supplementary to quantitative data.

Lived experience provides unique insights into the real-world impact and implementation of interventions, highlighting issues that may not be captured in quantitative studies. For example, qualitative research can reveal barriers to care experienced by marginalised groups (183), informing inclusive and effective guidelines (130). Recognising the value of lived experience can enhance the relevance and applicability of guidelines (184).

NICE routinely have two lay members on their guideline committees. On occasion, if lived experience is considered to be particularly pivotal, there can be more. Lay-members come from a range of backgrounds. They can be people with lived experience of a condition or disease (usually longer-term conditions), they can be the parent or carer of someone who had the condition, or on occasion they can be employees from grassroots third sector organisations who have access to many people who have used services and can relay their stories. As discussed in the previous chapter, there was often an assumption that lay-members on committees would engage better with QES than with quantitative data and a sense that, broadly speaking, QES and lay-members were more or less the same thing. This was presumably in the sense that the QES talks about people's lived experiences, and the purpose of lay members on committees is to reflect their own and others lived experience. Notably these views came from non-lay-member participants. Lay-members were clear that the breadth of opinion and insight into other people views through QES gave them a broader understanding than just their own. This chapter has already discussed the view that QES might not add anything to the lived experience of the lay members. It is an interesting observation that a sample size of two lay members is considered by some committee members to be broadly equivalent to a robustly sampled QES; and this within the context of a broader concern among committee members about the qualitative research sample sizes and their generalisability. Clearly it is not possible to have a representative sample of lay-members on a NICE committee, and the People and Communities team at NICE (who recruit

and support lay-members) do not aim to, or claim to, do this. A well-conducted QES including several well-sampled qualitative primary studies is far more likely to represent a meaningful breadth of views.

CHAPTER 7: SUMMARY AND RECOMMENDATIONS

Introduction

Taken together, the three studies in this thesis provide a layered understanding of the role of QES in guideline development. The systematic review (Chapter 2) outlined best practices and methodological ideals; the content analysis (Chapter 3) revealed how these are (or are not) implemented in NICE guidelines; and the qualitative study (Chapter 5) explored how QES are perceived and experienced by those involved in guideline development. The integration of these findings highlights a clear gap between methodological guidance and the problems of real-world application, and suggests that improving training, reporting standards, and the strategic use of QES could enhance their impact on guideline development.

In the spirit of Karl Marx, whose gravestone famously states “The philosophers have only interpreted the world, in various ways. The point, however, is to change it.” (from his Eleven Theses on Feuerbach, published in 1938), this conclusion will focus mostly on ‘changing it’ based on the preceding chapters that seek to ‘interpret it’. The chapter makes a series of practical suggestions that guideline producers might use to optimise the usefulness of QES to guideline committees who consider qualitative evidence as part of their evidence base. The chapter also discusses limitations of this thesis and the studies in it.

Optimising the use of QES in guideline development

Increasing Trust in the Credibility and Robustness of QES

Findings from this study suggest that there can be limited understanding of the qualitative paradigm amongst committee members and technical staff and, as a result, it is difficult for them to judge QES on its own merits. Committees that are going to consider QES as part of the evidence base for a guideline would benefit from a better understanding of pluralist research methods and their interpretation. This study suggests that training should include the formulation of qualitative review questions and the execution of QES. A more aggregative approach to synthesis might better suit both committee concerns about transparency, and the lack of qualitative background in most technical staff because these approaches are simpler (because they do not require an interpretive phase) and because it

is easier to see the mechanism for individual themes from papers contribution to QES findings.

The content analysis reported in Chapter 3 showed that NICE QES were good at meeting the reporting criteria for QES that mimicked those of quantitative systematic reviews but were less good at demonstrating those reporting criteria that were specific to qualitative methods. This could be interpreted as an indication that NICE technical staff are unfamiliar with both QES and qualitative research generally, a view upheld by the analysis of the primary qualitative study in Chapter 5.

The repeated references to difficulties in understanding QES methods and the need for additional support in Chapter 5 indicate significant educational gaps. Both committee members and technical staff would benefit from targeted training to better understand and conduct QES. This insight points to a need for ongoing professional development and capacity building within guideline development teams. Findings from the systematic review reported in the Chapter 2 section on training echo this acknowledgement that guideline developers are principally systematic reviewers who may have no background or expertise in qualitative methods, and therefore need training to be able to produce high-quality QES (106, 109, 173). It also may explain the gaps in reporting set out in Chapter 3 and discussed in the previous paragraph

Training programmes should focus on the principles and methods of qualitative research, the interpretation of qualitative data, and the integration of qualitative and quantitative evidence. By enhancing the skills and knowledge of committee members and technical staff, organisations can improve the quality and credibility of their guideline development processes.

Clearer reporting of QES in a way that makes sense to committees, for example using a template developed with input from committee members and chairs, and conforming to QES reporting standards alongside clearer description of the evidence and analysis might lead to greater understanding of the QES among committee members, and as a result, greater confidence in QES findings. This in turn might reduce the burden on technical staff re-presenting evidence at committee meetings that the committee would already be familiar with if the QES were clearly written up. Additionally, use of visualisations in QES

reports could prove useful, simpler examples might include word clouds, but more complex methods such as theme maps, conceptual models or frameworks might prove useful.

Presenting Findings

Committees want a high-level of detail from QES to compensate for their lack of familiarity with the evidence and methods. Additionally, QES can be quite unfocussed and contain a large volume of studies spread across a broad range of themes. This is a challenge for a technical team who are presenting the work because they cannot keep the presentation short, focussed, and concise.

The order in which evidence is reviewed and presented to committees has an impact on the perceived status of the evidence. Decisions about which evidence is core to each review question should inform the order of data presentation when qualitative and quantitative data are both relevant to a question should be based on which data are most relevant to the question and should be informed by current best practice in the form of mixed-methods approaches to convergent reviews and in terms of formal Evidence-to-Decision processes.

Committees, partly as a result of unfamiliarity with the methods of QES, have concerns about the derivation of themes in QES, and furthermore, about what those themes mean in the contexts in which they are trying to apply them in their 'recommendation-making'. They are concerned that they cannot always easily see the links between themes reported in the included studies and the overall themes of the QES. Adoption of an approach to QES where there is a clearer linkage between primary research themes and QES themes would be a valuable tool for generating a higher-level of trust in the findings of QES (see previous section). Appropriate methods would be those where it is possible to 'track' the way themes from primary studies are drawn together to form higher-level interpretations, for example through aggregative methods, or using a framework-based approach such as best-fit framework synthesis (185).

Using an established method of QES where there are clear links between the themes drawn from the primary studies and the final QES as described above would allow committees to see more clearly how the themes used in the final QES were derived, and this would have a double impact of the committee feeling less like they were putting 'blind trust' in the technical team, an idea with which they were uncomfortable (even though they seemed

willing to trust technical staff in relation to complex quantitative analyses such as Network Meta-Analyses (NMAs) and meta-regression). The size and complexity of some QES also make the QES documents and the committee presentations dense and wide-ranging. Clearer, focussed QES may reduce the uncertainty felt by committee members.

Not all guideline developers use formal Evidence-to-Decision processes such as those created by the GRADE working group (129) or by WHO-INTEGRATE (186) (for example, NICE do not use them). Instead, broader evidence is introduced by committee members and lay-members based on their expertise and experience and on a mandatory equalities and health inequality impact assessment that is completed at each stage of the guideline. This unstructured way of introducing evidence outside of the core effectiveness evidence that is created through systematic reviewing means that guideline committees can neglect important considerations and criteria or give undue weight to effectiveness evidence. The process of committee decision making, including all of their contextual considerations relating to resource costs, feasibility, acceptability, impact on health inequalities and so forth are described narratively to clarify how the committee made any decisions about recommendations.

Both the WHO-INTEGRATE and GRADE Evidence-to-Decision (EtD) frameworks have fields for evidence related to acceptability, feasibility, and equity, which in large part will encompass the outcomes of a QES (depending on the review question chosen for the synthesis), and which therefore creates a valid and equal place for QES alongside quantitative effectiveness evidence in the framework. Implementation of an EtD framework could address in large part all three overarching themes arising from the study in Chapters 4-6. This is because it demonstrates the value and credibility of QES and its place in the EtD process; shows the relative value of QES by putting it on a proportionate footing with quantitative evidence; and provides a clear framework for presenting data to the committee to inform their decision making. In turn, undertaking QES may benefit EtD frameworks by introducing potentially intellectually rigorous and rich data into the framework to replace speculative ideas about acceptability, equality etc that might be gleaned from surveys, single reports, or lay participant contributions only.

Value

Careful consideration of the usefulness of QES for each guideline area will maximise the utility of the QES. Committees are unlikely to value QES if it does not have a direct role in their recommendation-making.

The ability of QES to give a voice to populations who are often excluded from quantitative research may help to focus the committee's concerns about potential inequalities in health and enable them to nuance recommendations to ensure that they do not increase inequalities in health, or potentially even reduce them.

During the scope development phase of guideline planning, QES should always be considered as one of the options for producing evidence for guideline committees. This does not imply that a QES should always be blindly undertaken, but that the potential usefulness of one should always be considered. For a 'what works?' type of efficacy question, QES is unlikely to contribute to recommendation-making (although it may be useful for broader effectiveness questions). However, for questions about how, when, where, and for or by whom services or interventions should be delivered, QES could be a primary source of evidence.

Interpretation and championing of QES might be a key role for lay-members on committees. They could point out where findings accord with their experience (including how it made them reflect on their own experience and if it has raised issues that are true but which had not occurred to them, as they prioritise other experiences), but also where they do not (might be different, might cover many more themes - because they are only two people after all) - and also where the participants in the QES are different from them - and therefore the findings might be more or less generalisable. QES are also potentially a valuable support and resource for lay-members on committees. Two lay members cannot hope to meaningfully represent the broad range of people living with a condition and their parents/carers, and insight from a QES might help them achieve that broader representation.

Recent changes in the NICE guideline operating model have led to the formation of 'topic suites' that focus on particular areas of health policy and practice. Currently they are cancer, cardiometabolic, women's health, and mental health. The purpose of the suites is to

perform multiple updates of the guidelines that fall under that topic. The topic suites would be ideal areas where QES could support lay members by exploring the perspectives of people living with the conditions included in the topic suite, to provide a better understanding of their views and their perceptions of the services they receive.

Limitations of this thesis

Methodological Limitations

Sampling issues

The systematic review of methods in Chapter 2 only included papers that were specific to QES in relation to producing health and social care guidelines. This introduces two key limitations: firstly, a large proportion of the data published about ways to undertake QES is published outside of the specific area of guideline production, and therefore that data could not be considered as part of the systematic review itself. However, reference to this literature in other parts of the thesis addresses this issue, although it remains an important consideration. Secondly, it is also possible that papers published in fields other than health and social care might also address issues of using QES to inform guideline committees. Date limitations on the searches for the systematic review are unlikely to have a large impact since interest in using QES in health and social care guidelines is a relatively modern phenomenon. However, this is potentially an issue in the content analysis work described in Chapter 3. Due to resource limitations and the complexity of obtaining the data, a 5-year window was selected to sample from; however NICE has published QES both before and after that time window. Additionally, for both the content analysis and the primary qualitative study, a decision was made to focus on NICE as an exemplar of a guideline-producing organisation, zooming in on a much narrower field than the systematic review. By drawing all participants from a single agency, the study is inherently biased toward NICE's specific policies, practices, and organisational culture. This raises the possibility that the experiences or viewpoints of individuals from other agencies, which may have different approaches to guideline development, might be different. However, there is no reason to assume that this is the case. Having the time and resources to include participants from multiple agencies would have meant that the findings could have highlighted differences or broader trends across guideline producers as a whole, although NICE represented both a

practical and meaningful choice – given it is a guideline producer with an international reputation.

Measurement tools

The content analysis in Chapter 3 used the ENTREQ reporting criteria for QES as the basis for the evaluation of the QES. Two limitations are important:

ENTREQ is a reporting standard and is not a quality assessment tool for QES, so it is important not to conflate a high number of ENTREQ criteria being met with a QES being a 'good quality' QES, but merely a 'well-reported' QES. The use of ENTREQ criteria was useful for the purposes of this thesis because the intention was to assess the completeness of the reports being presented to the committee as a preparation for the primary qualitative study. Secondly, as previously described, although at the time of undertaking the content analysis, ENTREQ was the accepted reporting standard, and its use was an obvious choice as it was the one reporting standard available and applicable to the QES being assessed in Chapter 3, the ENTREQ standard is currently receiving negative attention because of potential issues (see Chapter 6 – challenges of presenting large volumes of data).

Data Limitations

Data availability/quality

Obtaining data from organisations methods manuals for the systematic review question was challenging, and for pragmatic reasons was restricted to those which were published either as academic publications in their own right or could be easily found by internet searching. A more comprehensive approach would have been to spend more time collating a list of organisations that undertake QES for the purposes of guideline production and to have contacted each of those agencies to ask for access to their QES methods documents.

In the content analysis, the data availability was somewhat patchy as described in that chapter (Chapter 3). NICE QES are not always written up as stand-alone documents but are often incorporated into a single review document alongside one or more quantitative reviews (with no formal mixed-methods synthesis, especially in the case of older guidelines). Also, many so-called QES reported including only one study (or no studies) and therefore were ineligible for inclusion. This means that data identification was not always possible in a

systematic way, leading to extensive hand-searching, which might lead to error. Double checking was not possible due to the resource limitations of a PhD.

For the primary qualitative study, conducting meetings online without video reduced the availability of visual cues that normally accompany qualitative interviewing. It was not possible for the interviewer and interviewees to interact with each other based on non-visual cues, and this may have led to some nuance being missed. Participants may be less communicative in online interviews if they feel less comfortable using this technology.

Additionally, as set out in the reflexivity section in Chapter 4, my status as an employee of the organisation could potentially have influenced my analysis and interpretation.

Data scope

In the qualitative study, using a 12-person purposive sample drawn from three guideline committees within a single agency presents several limitations that are important to acknowledge:

- The focus on three committees from one organisation means that the findings might not be easily transferable to other guideline committees, either within NICE or across other organisations. The participants, while selected to be as diverse as possible, represent only a subset of committee members within one agency, and their perspectives may not reflect those of individuals from other committees or from different agencies. Thus, the results might not be applicable to a broader range of committees or of other guideline developers.
- Since the sample is purposively selected from only three committees within a single agency, the findings are shaped by the experiences and views of those particular committees. Each committee, and each topic, had its own dynamics, challenges, and practices that may not be representative of others within NICE or from other agencies. The unique factors influencing these committees mean the findings could be skewed toward those specific experiences, limiting the overall transferability of the findings.

- Even though purposive sampling is intended to facilitate the inclusion of diverse professions and expertise, there may still be a lack of diversity within the perspectives of committee members and NICE staff, and this could result in a narrow range of insights. The lack of diversity in the sample could have affected the richness of the data and lead to findings that are more homogeneous than might be found in a varied sample.

Practical Constraints

Resources

As a self-funded, part-time PhD, no resources were available beyond those of the author.

Future research

As with any research project, the wish-list of things that could have been done is almost endless, however, several possibilities stand out in relation to the work described in the previous chapters.

Firstly, the systematic review in Chapter 2 was undertaken in the early part of the PhD and therefore is several years old already. Given the pace of change in QES over the past few years, an update to the systematic review in a year or so would be valuable, especially if resourced to identify methods documents from other guideline-producing bodies that may have been missed through a search of published literature.

Secondly, in the spirit of action research methods, revisiting the primary qualitative study and the content analysis after NICE has implemented any changes to methods of QES (should it do so in response to the findings of this thesis) would demonstrate the impact of those changes on the ways that qualitative evidence is (or isn't) used by committees to make recommendations when the methods are optimised.

Thirdly, the results of the primary qualitative study support ideas about the value of QES for representing stakeholder views (see Chapter 1) and also that they are a valuable tool for lay-members on NICE committees to introduce their own views and experiences to the committee. NICE have been slow to embrace the involvement of lay-people in the development of evidence syntheses, and a project to co-produce QES with committee lay-

members might be a valuable way to increase their role and visibility on guideline committees.

Fourthly, it is the nature of a PhD thesis to 'zoom in' to a very specific circumstance, in the case of this thesis the focus has primarily been on NICE. The narrow focus does not enable the author to understand whether the opportunities and challenges identified are specific to NICE, or whether they are similar to the opportunities and challenges faced by other guideline bodies who are using QES in their evidentiary process. It seems plausible that other guideline-producing bodies, who we know are using similar methods for integrating QES into their evidence base (as set out in Chapter 2), will be facing similar challenges.

CHAPTER 8: CONCLUSION

This PhD undertook a series of investigations intended to explore the role of QES in the development of health (including public health) and social care guidelines, and to consider how its role might be optimised.

To address the overall question, four sub-questions were proposed:

1. What methods and processes have been developed or proposed for incorporating QES into health guideline development?

2. How frequently were QES undertaken in the context of guideline development by an exemplar organisation (NICE) in the UK over a 5-year period, and how closely do the reports of those syntheses adhere to established reporting standards?

3. How do guideline committees at NICE use qualitative evidence from QES to inform recommendation-making, and what might be learned about best practice and future developments?

4. What are the views and perceptions of technical staff, committee experts and committee lay-members regarding how a QES contributes to committee discussions, and to the process of making recommendations?

The first question about methods and processes was primarily answered by the undertaking of a systematic review of the methodological literature (Chapter 2). The review showed a good deal of methodological congruence in the included literature about methods to undertake QES for application in guideline development but identified several gaps.

Two of the key gaps identified by Chapter 2 are that:

1. there was a notable lack of discussion regarding the role of guideline committees in interpreting evidence generated by QES and using it to develop recommendations. This omission is likely due to a general lack of understanding about the processes that committees undertake to use evidence for generating recommendations.

2. there is little contained in the included papers to explain or clarify the process of using QES to inform recommendations alongside other types of evidence and their own expertise and experience.

Both of these gaps in the published literature were incorporated into the planning and analysis of the primary qualitative study, which addresses them in its analysis, by exploring how committee members use qualitative evidence from QES to inform recommendation-making (question 3 above) and the views of both technical staff and committee members on how a QES contributes to committee discussions, and to the process of making recommendations (question 4 above). The analysis and discussion offer a perspective on these gaps and offer potential strategies to minimise concerns.

The second question was addressed by undertaking a quantitative content analysis that measured NICE's (as an exemplar guideline producer) use of QES in guideline production and sought to understand some patterns in the data. The study concluded that QES were under-used (very few comparatively were undertaken), and that the clearest factor in the apparent quality of a QES seems to have been the team that undertook it. Teams which produce well-reported QES seem to do so consistently, perhaps because they have staff with a particular interest or skill in this area. Solutions to this might include ensuring that staff undertaking QES have appropriate skills and supervision and providing clearer guidance about how a QES should be undertaken in terms of methods and processes. This is also borne out by the primary qualitative study, which identifies need for training (both of technical staff and committee members) in the methods and processes of QES, and their interpretation.

As well as addressing the gaps identified by the first two studies, the primary qualitative study reported in Chapters 4-6 provides a perspective on questions 3 and 4 by exploring technical staff and committee members' understandings of how QES was used in their committee experience and how it informed recommendations. There was clear uncertainty over how QES could and should be presented to a guideline committee, and how it could or should be used, particularly in relation to the quantitative evidence, when discussing and producing recommendations.

The overall conclusion is that QES unquestionably has a recognised role in guideline development. The findings of the studies contained within this thesis have contributed to an understanding of the current and potential role and use of QES in supporting guideline development. However, as this is a rapidly developing field, the role of QES often still remains ill-defined and insecure. If its place is to become fully established, unquestioned, and secure, further work is needed to build on the findings outlined in this thesis, specifically within guideline committees, to enhance a broader understanding of this evidence and its presentation.

CHAPTER 9: REFERENCES

1. Dwyer SC, Buckle JL. The Space Between: On Being an Insider-Outsider in Qualitative Research. International Journal of Qualitative Methods. 2009. 8(1):54-63. Available from: 10.1177/160940690900800105.

2. Berkovic D, Ayton D, Briggs AM, Ackerman IN. The View From the Inside: Positionality and Insider Research. International Journal of Qualitative Methods. 2020. 19:1609406919900828. Available from: 10.1177/1609406919900828.

3. Mercer J. The challenges of insider research in educational institutions: wielding a double-edged sword and resolving delicate dilemmas. Oxford Review of Education. 2007. 33(1):1-17. Available from: 10.1080/03054980601094651.

4. Yadav D. Criteria for Good Qualitative Research: A Comprehensive Review. The Asia-Pacific Education Researcher. 2021. 31(6):679-89. Available from: 10.1007/s40299-021-00619-0.

5. Nickerson RS. Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. Review of General Psychology. 1998. 2(2):175-220. Available from: 10.1037/1089-2680.2.2.175.

6. Brannick T, Coghlan D. In Defense of Being “Native”: The Case for Insider Academic Research. Organizational Research Methods. 2007. 10(1):59-74. Available from: 10.1177/1094428106289253.

7. Berger R. Now I see it, now I don’t: researcher’s position and reflexivity in qualitative research. Qualitative Research. 2013. 15(2):219-34. Available from: 10.1177/1468794112468475.

8. Finlay L. Negotiating the swamp: the opportunity and challenge of reflexivity in research practice. Qualitative Research. 2002. 2(2):209-30. Available from: 10.1177/146879410200200205.

9. Lincoln Y, Guba E Naturalistic Inquiry: Sage; 1985.

10. Graham R, Mancher M, Miller Wolman D, Greenfield S, Steinberg E. Clinical Practice Guidelines We Can Trust. Washington DC, US: Institute of Medicine (US) Committee on Standards for Developing Trustworthy Clinical Practice Guidelines Press NA. 2011.

11. Booth A, Sutton A, Papaioannou D Systematic Approaches to a Successful Literature Review. 2nd Edition ed. London, UK: Sage; 2016.

12. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ. 1996. 312(7023):71-2. Available from: 10.1136/bmj.312.7023.71.

13. Wouterse B, van Baal P, Versteegh M, Brouwer W. The Value of Health in a Cost-Effectiveness Analysis: Theory Versus Practice. Pharmacoeconomics. 2023. 41(6):607-17. Available from: 10.1007/s40273-023-01265-8.

14. IOM (Institute of Medicine) Clinical Practice Guidelines We Can Trust. Graham R, Mancher M, Miller Wolman D, Greenfield S, Steinberg E, editors. Washington (DC): The National Academies Press; 2011.

15. Cochrane A. Effectiveness and efficiency: Random reflections on health services. London, UK: Trust N. 1972.

16. Evidence-Based Medicine Working G. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA. 1992. 268(17):2420-5. Available from: 10.1001/jama.1992.03490170092032.

17. Greenhalgh T, Howick J, Maskrey N, Evidence Based Medicine Renaissance G. Evidence based medicine: a movement in crisis? BMJ. 2014. 348:g3725. Available from: 10.1136/bmj.g3725.

18. Grotz J, Ledgard M, Poland F Patient and Public Involvement in Health and Social Care Research: Palgrave Macmillan; 2020.

19. Williams O, Robert G, Martin GP, Hanna E, O'Hara J. Is Co-production Just Really Good PPI? Making Sense of Patient and Public Involvement and Co-production Networks. In: Bevir M, Waring J, editors. Decentring Health and Care Networks. Organizational Behaviour in Healthcare. Cham: Springer International Publishing; 2020. p. 213-37.

20. Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A. Synthesising qualitative and quantitative evidence: a review of possible methods. J Health Serv Res Policy. 2005. 10(1):45-53. Available from: 10.1177/135581960501000110.

21. Petticrew M, Roberts H. Evidence, hierarchies, and typologies: horses for courses. J Epidemiol Community Health. 2003. 57(7):527-9. Available from: 10.1136/jech.57.7.527.

22. Department of Health and Social Care Strengths-based approach: Practice Framework and Practice Handbook. London: Social Care Institute for Excellence; 2019.

23. Jordan Z, Lockwood C, Aromataris E, Munn Z The updated JBI model for evidence-based healthcare. Adelaide: Joanna Briggs Institute; 2016.

24. Kiyimba N, Lester JN, O'Reilly M. Examining Evidence-Based Practice and Practice-Based Evidence. Using Naturally Occurring Data in Qualitative Health Research. Cham: Springer International Publishing; 2019. p. 35-60.

25. Loder E, Groves T, Schroter S, Merino JG, Weber W. Qualitative research and The BMJ. BMJ. 2016. 352:i641. Available from: 10.1136/bmj.i641.

26. Saunders CH, Sierpe A, von Plessen C, Kennedy AM, Leviton LC, Bernstein SL, et al. Practical thematic analysis: a guide for multidisciplinary health services research teams engaging in qualitative analysis. BMJ. 2023. 381:e074256. Available from: 10.1136/bmj-2022-074256.

27. Purssell E, McCrae N. Critical Appraisal: Assessing the Quality of Evidence. How to Perform a Systematic Literature Review. Cham: Springer Nature Switzerland; 2024. p. 51-69.

28. Dawson AJ. Meta-synthesis of Qualitative Research. In: Liamputtong P, editor. Handbook of Research Methods in Health Social Sciences. Singapore: Springer Singapore; 2019. p. 785-804.

29. Institute of Medicine Clinical Practice Guidelines: Directions for a New Program. Field MJ, Lohr KN, editors. Washington, DC: The National Academies Press; 1990. 168 p.

30. National Institute for Health and Care Excellence (NICE) Developing NICE guidelines: the manual. Manchester: NICE; 2024.

31. Hempel S, Bolshakova M, Turner BJ, Dinalo J, Rose D, Motala A, et al. Evidence-Based Quality Improvement: a Scoping Review of the Literature. J Gen Intern Med. 2022. 37(16):4257-67. Available from: 10.1007/s11606-022-07602-5.

32. Straus S, Glasziou P, Richardson W, Haynes R Evidence-based medicine: How to practice and teach EBM. 5th edition ed. Edinburgh, UK: Elsevier; 2019.

33. Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. BMJ. 1999. 318(7182):527-30. Available from: 10.1136/bmj.318.7182.527.

34. Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. Evid Based Med. 2016. 21(4):125-7. Available from: 10.1136/ebmed-2016-110401.

35. Song F, Hooper, Loke Y. Publication bias: what is it? How do we measure it? How do we avoid it? Open Access Journal of Clinical Trials. 2013. 5:71-81. Available from: 10.2147/oajct.S34419.

36. Schulzke S. Assessment of Publication Bias. In: Patole S, editor. Principles and Practice of Systematic Reviews and Meta-Analysis. Cham: Springer International Publishing; 2021. p. 57-63.

37. Hopewell S, McDonald S, Clarke M, Egger M. Grey literature in meta-analyses of randomized trials of health care interventions. Cochrane Database Syst Rev. 2007. 2007(2):MR000010. Available from: 10.1002/14651858.MR000010.pub3.

38. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet. 2005. 365(9453):82-93. Available from: 10.1016/S0140-6736(04)17670-8.

39. Van Spall HG, Toren A, Kiss A, Fowler RA. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. JAMA. 2007. 297(11):1233-40. Available from: 10.1001/jama.297.11.1233.

40. Ioannidis JP. Evolution and translation of research findings: from bench to where? PLoS Clin Trials. 2006. 1(7):e36. Available from: 10.1371/journal.pctr.0010036.

41. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. BMJ. 2007. 334(7589):349-51. Available from: 10.1136/bmj.39070.527986.68.

42. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. Res Synth Methods. 2012. 3(2):80-97. Available from: 10.1002/jrsm.1037.

43. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. BMJ. 1994. 309(6965):1351-5. Available from: 10.1136/bmj.309.6965.1351.

44. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002. 21(11):1539-58. Available from: 10.1002/sim.1186.

45. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ. 2008. 337:a1655. Available from: 10.1136/bmj.a1655.

46. Petticrew M, Knai C, Thomas J, Rehfuess EA, Noyes J, Gerhardus A, et al. Implications of a complexity perspective for systematic reviews and guideline development in health decision making. BMJ Glob Health. 2019. 4(Suppl 1):e000899. Available from: 10.1136/bmjgh-2018-000899.

47. Moore GF, Evans RE, Hawkins J, Littlecott H, Melendez-Torres GJ, Bonell C, et al. From complex social interventions to interventions in complex social systems: Future directions and unresolved questions for intervention development and evaluation. Evaluation (Lond). 2019. 25(1):23-45. Available from: 10.1177/1356389018803219.

48. Petticrew M. When are complex interventions 'complex'? When are simple interventions 'simple'? Eur J Public Health. 2011. 21(4):397-8. Available from: 10.1093/eurpub/ckr084.

49. Hawe P, Shiell A, Riley T. Complex interventions: how "out of control" can a randomised controlled trial be? BMJ. 2004. 328(7455):1561-3. Available from: 10.1136/bmj.328.7455.1561.

50. Anderson R, Hardwick R, Pearson M, Byng R. Using Realist Approaches to Explain the Costs and Cost-Effectiveness of Programmes. In: Emmel N, Greenhalgh J, Manzano A, Monaghan M, editors. Using Realist Approaches to Explain the Costs and Cost-Effectiveness of Programmes: Sage; 2018.

51. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ. 2021. 374:n2061. Available from: 10.1136/bmj.n2061.

52. Noyes J, Booth A, Moore G, Flemming K, Tuncalp O, Shakibazadeh E. Synthesising quantitative and qualitative evidence to inform guidelines on complex interventions: clarifying the purposes, designs and outlining some methods. BMJ Glob Health. 2019. 4(Suppl 1):e000893. Available from: 10.1136/bmjgh-2018-000893.

53. Turner RM, Bird SM, Higgins JP. The impact of study size on meta-analyses: examination of underpowered studies in Cochrane reviews. PLoS One. 2013. 8(3):e59202. Available from: 10.1371/journal.pone.0059202.

54. Dixon-Woods M, Booth A, Sutton AJ. Synthesizing qualitative research: a review of published reports. Qualitative Research. 2007. 7(3):375-422. Available from: 10.1177/1468794107078517.

55. Walsh D, Downe S. Meta-synthesis method for qualitative research: a literature review. J Adv Nurs. 2005. 50(2):204-11. Available from: 10.1111/j.1365-2648.2005.03380.x.

56. Gough D, Thomas J, Oliver S. Clarifying differences between reviews within evidence ecosystems. Syst Rev. 2019. 8(1):170. Available from: 10.1186/s13643-019-1089-2.

57. Fletcher AJ. Applying critical realism in qualitative research: methodology meets method. International Journal of Social Research Methodology. 2016. 20(2):181-94. Available from: 10.1080/13645579.2016.1144401.

58. Wong G, Greenhalgh T, Westhorp G, Buckingham J, Pawson R. RAMESES publication standards: realist syntheses. BMC Med. 2013. 11:21. Available from: 10.1186/1741-7015-11-21.

59. Pawson R The science of evaluation: a realist manifesto. London: Sage; 2013.

60. Noblit GW, Hare RD *Meta-ethnography: Synthesizing qualitative studies*. Newbury Park, Calif: Sage Publications.; 1988.

61. Sandelowski M, Barroso J Handbook for synthesizing qualitative research: springer publishing company; 2006.

62. Barnett-Page E, Thomas J. Methods for the synthesis of qualitative research: a critical review. BMC Med Res Methodol. 2009. 9(1):59. Available from: 10.1186/1471-2288-9-59.

63. Booth A, Noyes J, Flemming K, Gerhardus A, Wahlster P, Van Der Wilt G, et al. Guidance on choosing qualitative evidence synthesis methods for use in health technology assessments of complex interventions. [Online]. Integrate - HTA; 2016. Available from: http://www.integrate-hta.eu/downloads/.

64. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. BMC Med Res Methodol. 2012. 12(1):181. Available from: 10.1186/1471-2288-12-181.

65. Hannes K, Lockwood C, Pearson A. A comparative analysis of three online appraisal instruments' ability to assess validity in qualitative research. Qual Health Res. 2010. 20(12):1736-43. Available from: 10.1177/1049732310378656.

66. Munthe-Kaas HM, Booth A, Sommer I, Cooper S, Garside R, Hannes K, et al. Developing CAMELOT for assessing methodological limitations of qualitative research for inclusion in qualitative evidence syntheses. Cochrane Evid Synth Methods. 2024. 2(6):e12058. Available from: 10.1002/cesm.12058.

67. France EF, Noyes J. Bangor University leads efforts to improve how research is reported within NHS care 2025 [accessed06/02/2025]. Available from: https://www.bangor.ac.uk/news/2025-01-28-bangor-university-leads-efforts-to-improve-how-research-is-reported-within-nhs-care.

68. France EF, Wells M, Lang H, Williams B. Why, when and how to update a meta-ethnography qualitative synthesis. Syst Rev. 2016. 5:44. Available from: 10.1186/s13643-016-0218-4.

69. Kavanagh J, Campbell F, Harden A, Thomas J. Mixed Methods Synthesis: A Worked Example. In: Hannes K, Lockwood C, editors. Synthesizing Qualitative Research. London: Wiley; 2011. p. 113-36.

70. Sutcliffe K, Harden A, Noyes J, Thomas J, Garside R. Integrating qualitative and quantitative review evidence. Draft version. In: Noyes J, Harden A, editors. Cochrane-Campbell Handbook for Qualitative Evidence Synthesis. London: Cochrane; 2024.

71. Lewin S, Booth A, Glenton C, Munthe-Kaas H, Rashidian A, Wainwright M, et al. Applying GRADE-CERQual to qualitative evidence synthesis findings: introduction to the series. Implement Sci. 2018. 13(Suppl 1):2. Available from: 10.1186/s13012-017-0688-3.

72. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q. 2004. 82(4):581-629. Available from: 10.1111/j.0887-378X.2004.00325.x.

73. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol. 2008. 8(1):45. Available from: 10.1186/1471-2288-8-45.

74. Carroll C, 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(2):149-54. Available from: 10.1002/jrsm.1128.

75. Hong QN, Pluye P, Bujold M, Wassef M. Convergent and sequential synthesis designs: implications for conducting and reporting systematic reviews of qualitative and quantitative evidence. Syst Rev. 2017. 6(1):61. Available from: 10.1186/s13643-017-0454-2.

76. O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S. Using text mining for study identification in systematic reviews: a systematic review of current approaches. Syst Rev. 2015. 4(1):5. Available from: 10.1186/2046-4053-4-5.

77. Downe S, Finlayson KW, Lawrie TA, Lewin SA, Glenton C, Rosenbaum S, et al. 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. 2019. 17(1):76. Available from: 10.1186/s12961-019-0467-5.

78. Glenton C, Lewin S, Lawrie TA, Barreix M, Downe S, Finlayson KW, et al. Qualitative Evidence Synthesis (QES) for Guidelines: Paper 3 – Using qualitative evidence syntheses to develop implementation considerations and inform implementation processes. Health Research Policy and Systems. 2019. 17(1):74. Available from: 10.1186/s12961-019-0450-1.

79. Dixon-Woods M, Cavers D, Agarwal S, Annandale E, Arthur A, Harvey J, et al. Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups. BMC Med Res Methodol. 2006. 6(1):35. Available from: 10.1186/1471-2288-6-35.

80. Barnett-Page E, Thomas J. Methods for the synthesis of qualitative research: a critical review. BMC Medical Research Methodology. 2009. 9(1):59. Available from: 10.1186/1471-2288-9-59.

81. Dekkers R, Carey L, Langhorne P. Methods for Qualitative Analysis and Synthesis. Making Literature Reviews Work: A Multidisciplinary Guide to Systematic Approaches. Cham: Springer International Publishing; 2022. p. 389-415.

82. Purssell E, McCrae N. Qualitative Reviews. How to Perform a Systematic Literature Review. Cham: Springer Nature Switzerland; 2024. p. 105-19.

83. Booth A, France E. Chapter 15 - Time-sensitive qualitative evidence syntheses. In: Noyes J, Harden A, editors. Cochrane-Campbell Handbook for Qualitative Evidence Synthesis, Version 1. London: Cochrane; 2024.

84. Harris J, Croot L, Thompson J, Springett J. How stakeholder participation can contribute to systematic reviews of complex interventions. J Epidemiol Community Health. 2016. 70(2):207-14. Available from: 10.1136/jech-2015-205701.

85. Ames HMR, France EF, Cooper S, Bianchim MS, Lewin S, Schmidt BM, et al. Assessing qualitative data richness and thickness: Development of an evidence-based tool for use in qualitative evidence synthesis. Cochrane Evid Synth Methods. 2024. 2(7):e12059. Available from: 10.1002/cesm.12059.

86. Noyes J, Harden A Cochrane-Campbell Handbook for Qualitative Evidence Synthesis: Cochrane; 2025.

87. Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist Synthesis: An Introduction. RMP Methods Paper 2/2004. 2004.

88. Siemens W, von Elm E, Binder H, Bohringer D, Eisele-Metzger A, Gartlehner G, et al. Opportunities, challenges and risks of using artificial intelligence for evidence synthesis. BMJ Evid Based Med. 2025.bmjebm-2024-113320. Available from: 10.1136/bmjebm-2024-113320.

89. O'Connor AM, Clark J, Thomas J, Spijker R, Kusa W, Walker VR, et al. Large language models, updates, and evaluation of automation tools for systematic reviews: a summary of significant discussions at the eighth meeting of the International Collaboration for the Automation of Systematic Reviews (ICASR). Syst Rev. 2024. 13(1):290. Available from: 10.1186/s13643-024-02666-2.

90. Gartlehner G, Kahwati L, Nussbaumer-Streit B, Crotty K, Hilscher R, Kugley S, et al. From promise to practice: challenges and pitfalls in the evaluation of large language models for data extraction in evidence synthesis. BMJ Evid Based Med. 2024.bmjebm-2024-113199. Available from: 10.1136/bmjebm-2024-113199.

91. Garritty C, Gartlehner G, Nussbaumer-Streit B, King VJ, Hamel C, Kamel C, et al. Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews. J Clin Epidemiol. 2021. 130:13-22. Available from: 10.1016/j.jclinepi.2020.10.007.

92. Shah W, Hillman T, Playford ED, Hishmeh L. Managing the long term effects of covid-19: summary of NICE, SIGN, and RCGP rapid guideline. BMJ. 2021. 372:n136. Available from: 10.1136/bmj.n136.

93. Cochrane. Cochrane COVID-19 Study Register 2020 [accessed 31/01/2025]. Available from: https://covid-19.cochrane.org/

94. Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ Open. 2017. 7(2):e012545. Available from: 10.1136/bmjopen-2016-012545.

95. Clark J, Glasziou P, Del Mar C, Bannach-Brown A, Stehlik P, Scott AM. A full systematic review was completed in 2 weeks using automation tools: a case study. J Clin Epidemiol. 2020. 121:81-90. Available from: 10.1016/j.jclinepi.2020.01.008.

96. Garritty C, Hamel C, Trivella M, Gartlehner G, Nussbaumer-Streit B, Devane D, et al. Updated recommendations for the Cochrane rapid review methods guidance for rapid reviews of effectiveness. BMJ. 2024. 384:e076335. Available from: 10.1136/bmj-2023-076335.

97. Epistemonikos Foundation, Megan Wainwright Consulting, Norwegian Institute of Public Health for the GRADE-CERQual Project Group. GRADE-CERQual Interactive Summary of Qualitative Findings (iSoQ). 1.0 ed. Oslo, Norway: Norwegian Institute of Public Health.

98. Critical Appraisal Skills Programme. CASP Qualitative Checklist 2018 [accessed 14/09/2019]. Available from: https://casp-uk.net/casp-tools-checklists/

99. Booth A, Sommer I, Noyes J, Houghton C, Campbell F, Cochrane Rapid Reviews Methods G, et al. Rapid reviews methods series: guidance on rapid qualitative evidence synthesis. BMJ Evid Based Med. 2024. 29(3):194-200. Available from: 10.1136/bmjebm-2023-112620.

100. Langlois EV, Tuncalp O, Norris SL, Askew I, Ghaffar A. Qualitative evidence to improve guidelines and health decision-making. Bull World Health Organ. 2018. 96(2):79-A. Available from: 10.2471/BLT.17.206540.

101. Glenton C, Lewin S, Lawrie TA, Barreix M, Downe S, Finlayson KW, et al. Qualitative Evidence Synthesis (QES) for Guidelines: Paper 3 - Using qualitative evidence syntheses to develop implementation considerations and inform implementation processes. Health Res Policy Syst. 2019. 17(1):74. Available from: 10.1186/s12961-019-0450-1.

102. Ring N, Jepson R, Ritchie K. Methods of synthesizing qualitative research studies for health technology assessment. Int J Technol Assess Health Care. 2011. 27(4):384-90. Available from: 10.1017/S0266462311000389.

103. Lewin S, Glenton C, Lawrie TA, Downe S, Finlayson KW, Rosenbaum S, et al. 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. 2019. 17(1):75. Available from: 10.1186/s12961-019-0468-4.

104. Carroll C. Qualitative evidence synthesis to improve implementation of clinical guidelines. BMJ. 2017. 356:j80. Available from: 10.1136/bmj.j80.

105. National Institute for Health and Care Excellence (NICE) Developing NICE guidelines: the manual. Manchester: NICE; 2018.

106. Glenton C, Lewin S, Norris SL. Using evidence from qualitative research to develop WHO guidelines (Chapter 15). World Health Organization Handbook for Guideline Development. 2nd ed. Geneva: WHO; 2016.

107. Hansen HP, Draborg E, Kristensen FB. Exploring qualitative research synthesis: the role of patients' perspectives in health policy design and decision making. Patient. 2011. 4(3):143-52. Available from: 10.2165/11539880-000000000-00000.

108. Kristensen FB, Sigmund H, editors. Health Technology Assessment Handbook Copenhagen: Danish Centre for Health Technology Assessment, National Board of Health 2007. [accessed 08/09/19] Available from: https://www.sst.dk/~/media/ECAAC5AA1D6943BEAC96907E03023E22.ashx.

109. Tan TP, Stokes T, Shaw EJ. Use of qualitative research as evidence in the clinical guideline program of the National Institute for Health and Clinical Excellence. Int J Evid Based Healthc. 2009. 7(3):169-72. Available from: 10.1111/j.1744-1609.2009.00135.x.

110. GRADE CERQual: Confidence in the Evidence from Reviews of Qualitative Research [accessed 14/09/2019]. Available from: https://www.cerqual.org/).

111. DECIDE collaboration. Evidence to Decision (EtD) framework [accessed 06/10/2019Available from: https://www.decide-collaboration.eu/evidence-decision-etd-framework

112. York Centre for Reviews and Dissemination (CRD). PROSPERO: International prospective register of systematic reviews [accessed 27/08/2019]. Available from: https://www.crd.york.ac.uk/prospero/.

113. Atkins S, Lewin S, Smith H, Engel M, Fretheim A, Volmink J. Conducting a meta-ethnography of qualitative literature: lessons learnt. BMC Med Res Methodol. 2008. 8:21. Available from: 10.1186/1471-2288-8-21.

114. Booth A. Searching for qualitative research for inclusion in systematic reviews: a structured methodological review. Syst Rev. 2016. 5:74. Available from: 10.1186/s13643-016-0249-x.

115. Booth A, Noyes J, Flemming K, Gerhardus A, Wahlster P, van der Wilt GJ, et al. Structured methodology review identified seven (RETREAT) criteria for selecting qualitative evidence synthesis approaches. J Clin Epidemiol. 2018. 99:41-52. Available from: 10.1016/j.jclinepi.2018.03.003.

116. Google. Google scholar [accessed 27/08/2019]. Available from: https://scholar.google.com/.

117. National Institute for Health and Care Excellence (NICE). EPPI Reviewer. 2018.

118. QSR International Pty Ltd. NVivo qualitative data analysis software. 12 ed2018.

119. Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) [accessed 14/09/2019]. Available from: www.prisma-statement.org.

120. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions. 6 ed: Cochrane; 2019.

121. Spencer L, Ritchie J, Lewis J, Dillon L. Quality in Qualitive Evaluation: A framework for assessing research evidence. 2003. [accessed 14/09/2019]. Available from: https://webarchive.nationalarchives.gov.uk/20140619055846/http:/www.civilservice.gov.uk/wp-content/uploads/2011/09/a_quality_framework_tcm6-38740.pdf..

122. Joanna Briggs Institute. QARI critical appraisal tool [accessed 14/09/2019]. Available from: https://joannabriggs.org/research/critical-appraisal-tools.html

123. ATLAS.ti [accessed 15/09/2019]. Available from: https://atlasti.com/.

124. MAXQDA [accessed 15/09/2019]. Available from: https://www.maxqda.com.

125. Joanna Briggs Institute. JBI SUMARI [accessed 15/09/2019]. Available from: http://jbisumari.org/.

126. France EF, Cunningham M, Ring N, Uny I, Duncan EAS, Jepson RG, et al. Improving reporting of meta-ethnography: the eMERGe reporting guidance. BMC Med Res Methodol. 2019. 19(1):25. Available from: 10.1186/s12874-018-0600-0.

127. Wong G, Greenhalgh T, Westhorp G, Pawson R. Development of methodological guidance, publication standards and training materials for realist and meta-narrative reviews: the RAMESES (Realist And Meta-narrative Evidence Syntheses - Evolving Standards) project. Health Serv Deliv Res. 2014. 2(30). Available from: 10.3310/hsdr02300.

128. Alonso-Coello P, Oxman AD, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines. BMJ. 2016. 353:i2089. Available from: 10.1136/bmj.i2089.

129. Alonso-Coello P, Schünemann HJ, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. BMJ. 2016. 353:i2016. Available from: 10.1136/bmj.i2016.

130. Lewin S, Glenton C. Are we entering a new era for qualitative research? Using qualitative evidence to support guidance and guideline development by the World Health Organization. Int J Equity Health. 2018. 17(1):126. Available from: 10.1186/s12939-018-0841-x.

131. Booth A. Acknowledging a Dual Heritage for Qualitative Evidence Synthesis: Harnessing the Qualitative Research and Systematic Review Research Traditions. : University of Sheffield.; 2013.

132. Booth A, Noyes J, Flemming K, Moore G, Tuncalp O, Shakibazadeh E. Formulating questions to explore complex interventions within qualitative evidence synthesis. BMJ Glob Health. 2019. 4(Suppl 1):e001107. Available from: 10.1136/bmjgh-2018-001107.

133. Ames H, Glenton C, Lewin S. Purposive sampling in a qualitative evidence synthesis: a worked example from a synthesis on parental perceptions of vaccination communication. BMC Med Res Methodol. 2019. 19(1):26. Available from: 10.1186/s12874-019-0665-4.

134. Braun V, Clarke V. To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. Qualitative Research in Sport, Exercise and Health. 2019. 13(2):201-16. Available from: 10.1080/2159676x.2019.1704846.

135. Munthe-Kaas HM, Glenton C, Booth A, Noyes J, Lewin S. 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):113. Available from: 10.1186/s12874-019-0728-6.

136. Alexander PE, Gionfriddo MR, Li SA, Bero L, Stoltzfus RJ, Neumann I, et al. A number of factors explain why WHO guideline developers make strong recommendations inconsistent with GRADE guidance. J Clin Epidemiol. 2016. 70:111-22. Available from: 10.1016/j.jclinepi.2015.09.006.

137. Noyes J, Booth A, Cargo M, Flemming K, Garside R, Hannes K, et al. Cochrane Qualitative and Implementation Methods Group guidance series-paper 1: introduction. J Clin Epidemiol. 2018. 97:35-8. Available from: 10.1016/j.jclinepi.2017.09.025.

138. Cochrane Qualitative and Implementation Methods Group. Cochrane Qualitative & Implementation Methods Group [accessed 14/10/2019]. Available from: https://methods.cochrane.org/qi/.

139. Bingham AJ. From Data Management to Actionable Findings: A Five-Phase Process of Qualitative Data Analysis. International Journal of Qualitative Methods. 2023. 22:16094069231183620. Available from: 10.1177/16094069231183620.

140. World Health Organization WHO handbook for guideline development. 2nd ed ed. Geneva: World Health Organization; 2014.

141. Wang YY, Liang DD, Lu C, Shi YX, Zhang J, Cao Y, et al. An exploration of how developers use qualitative evidence: content analysis and critical appraisal of guidelines. BMC Med Res Methodol. 2020. 20(1):160. Available from: 10.1186/s12874-020-01041-8.

142. Carmona C, Baxter S, Carroll C. Systematic review of the methodological literature for integrating qualitative evidence syntheses into health guideline development. Res Synth Methods. 2021. 12(4):491-505. Available from: 10.1002/jrsm.1483.

143. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res. 2005. 15(9):1277-88. Available from: 10.1177/1049732305276687.

144. Berelson B Content analysis in communication research. New York, NY, US: Free Press; 1952. 220- p.

145. Bengtsson M. How to plan and perform a qualitative study using content analysis. NursingPlus Open. 2016. 2:8-14. Available from: 10.1016/j.npls.2016.01.001.

146. Flemming K, Noyes J. Qualitative Evidence Synthesis: Where Are We at? International Journal of Qualitative Methods. 2021. 20. Available from: 10.1177/1609406921993276.

147. Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2023.

148. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. Journal of Open Source Software. 2019. 4(43). Available from: 10.21105/joss.01686.

149. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2016.

150. National Institute for Health and Care Excellence (NICE). Workplace health: management practices. 2016.

151. National Institute for Health and Care Excellence (NICE). Attention deficit hyperactivity disorder: diagnosis and management. 2019.

152. National Institute for Health and Care Excellence (NICE). Cataracts in adults: management. 2017.

153. Ramanadhan S, Revette AC, Lee RM, Aveling EL. Pragmatic approaches to analyzing qualitative data for implementation science: an introduction. Implement Sci Commun. 2021. 2(1):70. Available from: 10.1186/s43058-021-00174-1.

154. Cooke A, Smith D, Booth A. Beyond PICO: the SPIDER tool for qualitative evidence synthesis. Qual Health Res. 2012. 22(10):1435-43. Available from: 10.1177/1049732312452938.

155. Uttley L, Montgomery P. The influence of the team in conducting a systematic review. Syst Rev. 2017. 6(1):149. Available from: 10.1186/s13643-017-0548-x.

156. Buus N, Perron A. The quality of quality criteria: Replicating the development of the Consolidated Criteria for Reporting Qualitative Research (COREQ). Int J Nurs Stud. 2020. 102:103452. Available from: 10.1016/j.ijnurstu.2019.103452.

157. Braun V, Clarke V. How do you solve a problem like COREQ? A critique of Tong et al.'s (2007) Consolidated Criteria for Reporting Qualitative Research. Methods in Psychology. 2024. 11:100155. Available from: 10.1016/j.metip.2024.100155.

158. National Institute for Health and Care Excellence (NICE). International strategy for NICE: 2022 to 2025 [accessed 30/12/2024]. Available from: https://nice-international.nice.org.uk/international-strategy/index.html.

159. National Institute for Health and Care Excellence (NICE). How we develop NICE guidelines [accessed 01/03/2024]. Available from: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-guidelines/how-we-develop-nice-guidelines.

160. National Institute for Health and Care Excellence (NICE). Adrenal insufficiency: identification and management. 2024.

161. Connelly J. Realism in evidence based medicine: interpreting the randomised controlled trial. J Health Organ Manag. 2004. 18(2-3):70-81. Available from: 10.1108/14777260410538861.

162. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology. 2006. 3(2):77-101. Available from: 10.1191/1478088706qp063oa.

163. Campbell S, Greenwood M, Prior S, Shearer T, Walkem K, Young S, et al. Purposive sampling: complex or simple? Research case examples. J Res Nurs. 2020. 25(8):652-61. Available from: 10.1177/1744987120927206.

164. Kallio H, Pietila AM, Johnson M, Kangasniemi M. Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. J Adv Nurs. 2016. 72(12):2954-65. Available from: 10.1111/jan.13031.

165. Braun V, Clarke V Thematic analysis: a practical guide. London, UK: Sage; 2021.

166. Barriball KL, While A. Collecting data using a semi-structured interview: a discussion paper. J Adv Nurs. 1994. 19(2):328-35. Available from: 10.1111/j.1365-2648.1994.tb01088.x.

167. Braun V, Clarke V Successful Qualitative Research: A Practical Guide for Beginners. London: Sage; 2013.

168. Tracy SJ. Qualitative Quality: Eight “Big-Tent” Criteria for Excellent Qualitative Research. Qualitative Inquiry. 2010. 16(10):837-51. Available from: 10.1177/1077800410383121.

169. Shenton AK. Strategies for ensuring trustworthiness in qualitative research projects. Education for Information. 2004. 22(2):63-75. Available from: 10.3233/efi-2004-22201.

170. Hadi MA, Jose Closs S. Ensuring rigour and trustworthiness of qualitative research in clinical pharmacy. Int J Clin Pharm. 2016. 38(3):641-6. Available from: 10.1007/s11096-015-0237-6.

171. Stahl N, King J. Expanding Approaches for Research: Understanding and Using Trustworthiness in Qualitative Research. 2020. 44:26-8.

172. Flemming K, Booth A, Garside R, Tuncalp O, Noyes J. Qualitative evidence synthesis for complex interventions and guideline development: clarification of the purpose, designs and relevant methods. BMJ Glob Health. 2019. 4(Suppl 1):e000882. Available from: 10.1136/bmjgh-2018-000882.

173. Knaapen L, Colvin CJ, Cowl J, van der Weijden T. How to include qualitative research on patient views in guidelines. GIN public toolkit. 2015.

174. Silveira Bianchim M, France EF, Noyes J. Using visual methods to further enhance qualitative evidence synthesis. Cochrane Evid Synth Methods. 2024. 2(12):e70009. Available from: 10.1002/cesm.70009.

175. Ames H, Booth A, Noyes J. Chapter 6: Selecting studies and sampling. In: Noyes J, Harden A, editors. Cochrane-Campbell Handbook for Qualitative Evidence Synthesis, Version 1. London: Cochrane; 2024.

176. Pope C, Ziebland S, Mays N. Analysing Qualitative Data. Qualitative Research in Health Care2006. p. 63-81.

177. Creswell J, Plano Clark V Designing and Conducting Mixed Methods Research. Thousand Oaks, CA: Sage; 2017.

178. The Cochrane Collaboration. Review Manager (RevMan). 5.4 ed2020.

179. LoBiondo-Wood G, Haber J Nursing Research: Methods and Critical Appraisal for Evidence-Based Practice. St. Louis: Mosby Elsevier; 2014.

180. Greenhalgh T, Thorne S, Malterud K. Time to challenge the spurious hierarchy of systematic over narrative reviews? Eur J Clin Invest. 2018. 48(6):e12931. Available from: 10.1111/eci.12931.

181. O'Cathain A, Murphy E, Nicholl J. Three techniques for integrating data in mixed methods studies. BMJ. 2010. 341:c4587. Available from: 10.1136/bmj.c4587.

182. Barry CA, Britten N, Barber N, Bradley C, Stevenson F. Using reflexivity to optimize teamwork in qualitative research. Qual Health Res. 1999. 9(1):26-44. Available from: 10.1177/104973299129121677.

183. Konno R, Inoue K, Matsushita Y, Hashimoto K, Wiechula R, To T, et al. Barriers to Advance Care Planning in Older Adults With Dementia, Their Families and Healthcare Professionals: An Umbrella Review of Qualitative Evidence. Res Aging. 2024. 46(5-6):339-58. Available from: 10.1177/01640275241227909.

184. Brett J, Staniszewska S, Mockford C, Herron-Marx S, Hughes J, Tysall C, et al. Mapping the impact of patient and public involvement on health and social care research: a systematic review. Health Expect. 2014. 17(5):637-50. Available from: 10.1111/j.1369-7625.2012.00795.x.

185. Carroll C, Booth A, Leaviss J, Rick J. "Best fit" framework synthesis: refining the method. BMC Med Res Methodol. 2013. 13(1):37. Available from: 10.1186/1471-2288-13-37.

186. Rehfuess EA, Stratil JM, Scheel IB, Portela A, Norris SL, Baltussen R. The WHO-INTEGRATE evidence to decision framework version 1.0: integrating WHO norms and values and a complexity perspective. BMJ Glob Health. 2019. 4(Suppl 1):e000844. Available from: 10.1136/bmjgh-2018-000844.

APPENDICES

APPENDIX A: REVIEW PROTOCOL

Table 7: Review protocol

Field	Content
Review title	A systematic review of the methodological literature for integrating qualitative evidence syntheses into guideline development
Review question	What methods and processes are proposed in the methodological literature for incorporating QES into evidence-based health guideline development?
Objective	To capture the published and key grey literature that describe methods for incorporating QES into evidence-based health guidelines. This structured methodological review will capture primary data from coding of relevant papers using a template approach (described below) focussing on key methodological elements of the process: • Question design • Searching • Sifting • Coding/data extraction • Data synthesis • Use of logic models/frameworks • Integration with quantitative data/reviews

	• Decision making processes
Searches	The following databases will be searched: • Embase • MEDLINE • CINAHL • PsycINFO See appendix B for full search strategies for all databases. Search coverage will be checked using sentinel papers. Other searches: • Google – first six pages • Reference list screening of all included papers • Citation searching of all included papers Grey Literature will be included if cited on a reference list at screening or identified through the Google search.

	If less than ten papers are identified by the methods above, supplemental searches will be conducted as follows: • Hand-searching of key journals • Opportunistic methods: - Websites of research groups involved with QES in health or guideline development - JISCmail QES list - Expert consultation The searches will be re-run 6 months before final submission and further papers retrieved for inclusion. The full search strategies for MEDLINE database and full detail of additional and opportunistic searches will be published in the final review.
Condition or domain being studied	Qualitative evidence synthesis in evidence-based health guidelines.
Target	**Inclusion**: Papers that describe or propose methods and processes for incorporating QES into different

	kinds of evidence-based health (clinical or public health) or social care guideline. **Exclusion**: • Papers that do not describe a clear, extractable methodology or process. • Qualitative evidence syntheses themselves.
Content	Formal methods for qualitative evidence synthesis as part of the process of producing an evidence-based guideline. For the purposes of this review a qualitative evidence synthesis is defined as: "a process of systematically and transparently combining evidence from individual qualitative studies to create new understanding by comparing and analysing concepts and findings from different sources of evidence with a focus on the same topic of interest" An evidence-based guideline is defined as: "A series of statements or recommendations intended to optimise care or to inform decisions made by providers and service users. At a minimum:

	• the process should include a comprehensive and systematic assessment of the best available evidence • the evidence should be subject to a transparent decision-making process • the process should produce statements or recommendations, based on interpretation of the evidence (including the strength and quality of the evidence) that intend to improve care or support health decision making.”
Comparator/Reference standard/Confounding factors	Not applicable
Types of study to be included	Papers, online sources, or monographs that describe or prescribe methods for incorporating QES into evidence-based health guidelines. Sources must give sufficient detail to allow extraction of different stages of the process and the methods used in those stages.
Other exclusion criteria	• Non-English language • Books

	• Theses
Context	Any
Primary outcomes (critical outcomes)	Description of the methods used to perform the QES and integrate it with other data to inform the decisions made, specifically: • Question design • Searching • Sifting • Coding/data extraction • Data synthesis • Use of logic models/frameworks • Integration with quantitative data/reviews • Decision making processes
Secondary outcomes (important outcomes)	• Description of obstacles and challenges • Optimal methods • Adverse outcomes • Other emergent themes

Data extraction (selection and coding)	All references identified by the searches and from other sources will be uploaded into EPPI reviewer 5 or EndNote and de-duplicated. The full text of potentially eligible papers will be retrieved and will be assessed in line with the criteria outlined above. Papers suitable for inclusion will be uploaded into NVivo 12 and coded as described below
Risk of bias (RoB) (quality) assessment	Formal RoB assessment is not applicable since these are not primary studies, however the author may comment on the design and key weaknesses of included papers.
Strategy for data synthesis	Data will be extracted by open coding in NVivo 12, using a template codeset that is designed to capture the different methodological stages of guideline development and the processes of integrating QES with other data to inform healthcare decision making in guidelines. Papers will also be coded for emergent themes to allow for unanticipated relevant detail to be captured. Further data analysis may take place outside of NVivo using Microsoft Office packages.
Analysis of sub-groups	If there is sufficient consistency of named approaches, codes will be grouped by approach to explore whether different approaches offer different benefits and difficulties.
	☐ Diagnostic

	☐	Prognostic
	☒	Qualitative
	☐	Epidemiologic
	☐	Service Delivery
	☐	Other (please specify)
Funding sources/sponsor	This review is being completed as part of a self-funded PhD	
Conflicts of interest	None	
Dissemination plans	This review will form part of a PhD thesis and may be written up for publication	
Keywords	Qualitative evidence synthesis, Clinical guidelines, Public health guidelines, Social care guidelines	
Details of existing review of same topic by same authors	None	

APPENDIX B: SYSTEMATIC SEARCH HISTORY

Name: Chris Carmona
Topic/question details: **What methods and processes are proposed in the methodological literature for incorporating QES into evidence-based health guideline development?**

Table 8: Summary of search results

Databases	**Date searched**	**Version/files**	**No. retrieved**
Embase (Ovid)	15/8/19	Embase <1996 to 2019 Week 32>	3727
MEDLINE (Ovid)	15/8/19	Ovid MEDLINE(R) <1996 to August 13, 2019>	719
MEDLINE In-Process (Ovid)	15/8/19	Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <1946 to August 13, 2019>	63
CINAHL(EBSCO)	15/8/19	CINAHL 1998 to 2019	884

PsycINFO (Ovid)	15/8/19	PsycINFO <2002 to August Week 1 2019>	410

Table 9: Search strategies - medline

Database: MEDLINE
Results: 719 Database: Ovid MEDLINE(R) <1996 to August 13, 2019> Search Strategy: --- 1 (("semi-structured" or semistructured or unstructured or informal or "in-depth" or indepth or "face-to-face" or structured or guide) adj2 (interview* or discussion* or questionnaire*)).tw,kw. (86169) 2 (focus group* or qualitative or ethnograph* or fieldwork or thematic or phenomenological or "grounded theory" or "field work" or "key informant").tw,kw. (192379) 3 interviews as topic/ or focus groups/ or narration/ or qualitative research/ (113878) 4 1 or 2 or 3 (283347) 5 ("meta-synthesis" or "Meta-synthesis" or "framework synthesis").tw,kw. (687) 6 exp Review Literature as Topic/ (12145) 7 (review* or overview*).ti. (304571)

8 (systematic* adj5 (review* or overview*)).tw. (125996)

9 (qualitative* adj5 (review* or overview* or synthesis)).tw. (5867)

10 (integrat* adj3 (research or review* or literature)).tw. (8723)

11 (pool* adj2 (analy* or data)).tw. (22510)

12 (handsearch* or (hand adj3 search*)).tw. (7737)

13 (manual* adj3 search*).tw. (4754)

14 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 (374646)

15 ((clinical or care*) adj3 pathway*).ti,ab,kw. (8488)

16 (practice adj3 parameter*).ti,ab,kw. (1186)

17 (guidance* or guideline*).ti. (64813)

18 algorithms/ or clinical protocols/ or critical pathway/ or guidelines as topic/ or practice guidelines as topic/

or Health Planning Guidelines/ or practice guideline/ (417898)

19 Technology Assessment, Biomedical/ (7283)

20 15 or 16 or 17 or 18 or 19 (448362)

21 4 and 14 and 20 (752)

22 limit 21 to english language (719)

Table 10: Search strategies - medline in-process

Database: MEDLINE in-Process

Results: 63

Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <1946 to August 13, 2019>

Search Strategy:

---

1 (("semi-structured" or semistructured or unstructured or informal or "in-depth" or indepth or "face-to-face" or

structured or guide) adj2 (interview* or discussion* or questionnaire*)).tw,kw. (15764)

2 (focus group* or qualitative or ethnograph* or fieldwork or thematic or phenomenological or "grounded theory" or

"field work" or "key informant").tw,kw. (42243)

3 interviews as topic/ or focus groups/ or narration/ or qualitative research/ (0)

4 1 or 2 or 3 (50295)

5 ("meta-synthesis" or "Meta-synthesis" or "framework synthesis").tw,kw. (204)

6 exp Review Literature as Topic/ (0)

7 (review* or overview*).ti. (77895)

8 (systematic* adj5 (review* or overview*)).tw. (30009)

9 (qualitative* adj5 (review* or overview* or synthesis)).tw. (1472)

10 (integrat* adj3 (research or review* or literature)).tw. (1866)

11 (pool* adj2 (analy* or data)).tw. (3849)

12 (handsearch* or (hand adj3 search*)).tw. (985)

13 (manual* adj3 search*).tw. (826)

14 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 (89196)

15 ((clinical or care*) adj3 pathway*).ti,ab,kw. (1682)

16 (practice adj3 parameter*).ti,ab,kw. (115)

17 (guidance* or guideline*).ti. (10149)

18 algorithms/ or clinical protocols/ or critical pathway/ or guidelines as topic/ or practice guidelines as topic/

or Health Planning Guidelines/ or practice guideline/ (38)

19 Technology Assessment, Biomedical/ (0)

20 15 or 16 or 17 or 18 or 19 (11920)

21 4 and 14 and 20 (64)

22 limit 21 to english language (63)

Table 11: Search strategies - embase

Database: Embase
Database: Embase <1996 to 2019 Week 32> Search Strategy:

---

1 (("semi-structured" or semistructured or unstructured or informal or "in-depth" or indepth or "face-to-face" or

structured or guide) adj2 (interview* or discussion* or questionnaire*)).tw,kw. (135921)

2 (focus group* or qualitative or ethnograph* or fieldwork or thematic or phenomenological or "grounded theory" or

"field work" or "key informant").tw,kw. (301375)

3 exp interview/ or narrative/ or qualitative research/ (291630)

4 or/1-3 (531073)

5 (("meta-synthesis" or "Meta-synthesis" or "framework synthesis" or thematic) adj2 analysis).tw,kw. (22818)

6 (review* or overview*).ti. (471369)

7 (systematic* adj5 (review* or overview*)).tw. (204724)

8 (qualitative* adj5 (review* or overview* or synthesis)).tw. (9247)

9 (integrat* adj3 (research or review* or literature)).tw. (12772)

10 (pool* adj2 (analy* or data)).tw. (40771)

11 (handsearch* or (hand adj3 search*)).tw. (10771)

12 (manual* adj3 search*).tw. (7033)

13 or/5-12 (603620)

14 ((clinical or care*) adj3 pathway*).ti,ab,kw. (18202)

15 (practice adj3 parameter*).ti,ab,kw. (1947)

16 (guidance* or guideline*).ti. (102791)

17 algorithm/ or clinical protocols/ or Health care planning/ or practice guideline/ (753032) 18 Biomedical Technology Assessment/ (11507) 19 or/15-18 (796687) 20 4 and 13 and 19 (3821) 21 limit 20 to english language (3727)

Table 12: Search strategies - CINAHL

Database: CINAHL
Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - CINAHL

#	Query	Results
S4	S1 AND S2 AND S3	884
S3	SYSTEMATIC REVIEW/ OR (review* OR overview*).ti,ab OR (systematic* ADJ5 (review* OR overview*)).ti,ab OR (qualitative* ADJ5 (review* OR overview* OR synthesis)).ti,ab OR (integrat* ADJ3 (research OR review* OR literature)).ti,ab OR (pool* ADJ2 (analy* OR data)).ti,ab OR (handsearch* OR (hand ADJ3 search*)).ti,ab OR (manual* ADJ3 search*).ti,ab	111,849

S2	((("semi-structured" OR semistructured OR unstructured OR informal OR "in-depth" OR indepth OR "face-to-face" OR structured OR guide) ADJ2 (interview* OR discussion* OR questionnaire*)).ti,ab OR (focus group* OR qualitative OR ethnograph* OR fieldwork OR thematic OR phenomenological OR "grounded theory" OR "field work" OR "key informant").ti,ab OR INTERVIEWS/ OR FOCUS GROUPS/ OR NARRATIVES/ OR QUALITATIVE STUDIES/ OR (("meta-synthesis" OR "Meta-synthesis" OR "framework synthesis" OR thematic) ADJ2 analysis).ti,ab)	336,371
S1	(((clinical OR care*) ADJ3 pathway*).ti,ab OR (practice ADJ3 parameter*).ti,ab OR (guidance* OR guideline*).ti,ab OR DECISION SUPPORT TECHNIQUES/ OR PROTOCOLS/ OR PRACTICE GUIDELINES/ OR (Health technology assessment).ti,ab	174,936

Table 13: Search strategies - PsycINFO

Database: PsycINFO
Database: PsycINFO <2002 to August Week 1 2019> Search Strategy: ----------   1 (("semi-structured" or semistructured or unstructured or informal or "in-depth" or indepth or "face-to-face" or structured or guide) adj2 (interview* or discussion* or questionnaire*)).ti,ab. (76500)

2 (group* or qualitative or ethnograph* or fieldwork or thematic or phenomenological).ti,ab. (672597)

3 "grounded theory".ti,ab. (13384)

4 "field work".ti,ab. (445)

5 "key informant".ti,ab. (1393)

6 Semi-structured interview/ or focus group interview/ or narrative analysis/ or qualitative methods/ (8863)

7 or/1-6 (708637)

8 (("meta-synthesis" or "Meta-synthesis" or "framework synthesis" or thematic) adj2 analysis).ti,ab. (12331)

9 exp Literature Review/ (1076)

10 (review* or overview*).ti. (112644)

11 (systematic* adj5 (review* or overview*)).ti,ab. (29820)

12 (qualitative* adj5 (review* or overview* or synthesis)).ti,ab. (3630)

13 (integrat* adj3 (research or review* or literature)).ti,ab. (8925)

14 (pool* adj2 (analy* or data)).ti,ab. (2906)

15 (handsearch* or (hand adj3 search*)).ti,ab. (1177)

16 (manual* adj3 search*).ti,ab. (905)

17 or/8-16 (144498)

18 ((clinical or care*) adj3 pathway*).ti,ab. (2022)

19 (practice adj3 parameter*).ti,ab. (339)

20 (guidance* or guideline*).ti. (7644)

21 treatment guidelines/ (6190) 22 or/18-21 (13726) 23 7 and 17 and 22 (445) 24 limit 23 to english language (410)

Table 14: Search strategies - supplementary

Supplementary Search Techniques:
Google – qualitative synthesis guideline development returned 16 relevant records on the first eight pages. Reference list searching of included papers to be undertaken.

APPENDIX C: STUDY INCLUSION CHECKLIST

Table 15: Study inclusion checklist

	Author	2000+	QES?	Guidelines?	Methods?	Include?	Refs
1	Booth 2016	Y	Y	Y	Y	Y	0
2	Campbell 2011	Y	Y	N	Y	N	NA
3	Carroll 2017	Y	Y	Y	Y	Y	0
4	Cowles 2017	Y	N	Y	Y	N	NA
5	DeJean 2016	Y	Y	N	N	N	NA
6	Dixon-Woods 2001	Y	Y	N	Y	N	NA
7	Downe 2019	Y	Y	Y	Y	Y	1
8	Eakin 2003	Y	Y	N	N	N	NA
9	Fadlallah 2019	Y	N	N	Y	N	NA
10	Flemming 2019	Y	Y	Y	Y	Y	1
11	Florez 2018	Y	N	Y	Y	N	NA

	Author	2000+	QES?	Guidelines?	Methods?	Include?	Refs
12	Gargon 2017	Y	N	Y	N	N	NA
13	Glenton 2016	Y	Y	Y	Y	Y	0
14	Glenton 2019	Y	Y	Y	Y	Y	0
15	Gould 2010	Y	Y	Y	Y	Y	0
16	Grant 2018	Y	N	Y	Y	N	NA
17	Grummer-Strawn 2018	Conference abstract				N	NA
18	Hansen 2011	Y	Y	Y	Y	Y	0
19	Harden 2017	Y	Y	N	Y	N	NA
20	Huls 2018	Conference abstract				N	NA
21	Kelson 2015	Y	N	Y	N	N	NA
22	Knaapen 2015	Y	Y	Y	Y	Y	1
23	Korhonen 2013	Y	Y	N	Y	N	NA
24	Krahn 2008	Y	N	Y	N	N	NA

	Author	2000+	QES?	Guidelines?	Methods?	Include?	Refs
25	Kristensen 2007	Y	Y	Y	Y	Y	0
26	Langlois 2018	Y	Y	Y	N	N	NA
27	Lewin 2015	Y	Y	Y	N	N	NA
28	Lewin 2018	Y	Y	Y	Y	Y	0
29	Lewin 2019	Y	Y	Y	Y	Y	0
30	Li 2015	Y	N	Y	Y	N	NA
31	Longworth 2011	Y	N	Y	N	N	NA
32	McPherson 2018	Y	N	Y	N	N	NA
33	Murphy 1998	N	N	Y	Y	N	NA
34	NICE Manual 2018	Y	Y	Y	Y	Y	0
35	Noyes 2019	Y	N	Y	Y	N	NA
36	Opiyo 2013	Y	N	Y	Y	N	NA
37	Pope 2002	Y	N	Y	Y	N	NA

	Author	2000+	QES?	Guidelines?	Methods?	Include?	Refs
38	Ring 2010	Y	Y	Y	Y	Y	1
39	Ring 2011	Y	Y	Y	Y	Y	0
40	Roddis 2018	Y	N	Y	N	N	NA
41	Roddis 2019	Y	Y	Y	N	N	NA
42	Rosedale 2012	Y	N	N	Y	N	NA
43	Saunders 2015	Y	N	N	N	N	NA
44	SBU qual methods Hbk	Y	Y	Y	Y	Y	0
45	Schünemann 2006	Y	N	Y	N	N	NA
46	Staniszwewska 2014	Y	N	Y	Y	N	NA
47	Sundberg 2017	Y	N	Y	N	N	NA
48	Tan 2009	Y	Y	Y	Y	Y	0
49	Tong 2014	Y	Y	N	Y	N	NA
50	Van Wesel 2014	Y	Y	N	Y	N	NA

	Author	2000+	QES?	Guidelines?	Methods?	Include?	Refs
51	Weich 2018	Y	N	Y	N	N	NA
52	Wieringa 2018	Y	N	Y	N	N	NA
53	Zhang 2017	Y	N	Y	N	N	NA

Key *[shaded papers are included in the analysis]*

2000+ - Is the paper published after 2000?

QES – does the paper discuss synthesis or review of qualitative evidence?

Guidelines – does the paper refer specifically to guideline development (or can it be assumed)?

Methods – is there extractable methodological detail?

APPENDIX D: PAPERS EXCLUDED AT FULL TEXT

Table 16: Papers excluded at full text

Study	Code [Reason]
Campbell R, Pound P, Morgan M et al. (2011) Evaluating meta-ethnography: systematic analysis and synthesis of qualitative research.. Health technology assessment (Winchester, England) 15(43): 1-164	- Not about guidelines *[Mentions usefulness to HTA but doesn'tdoesn't actually discuss this.]*
Cowles, Emma, Marsden, Grace, Cole, Amanda et al. (2017) A Review of NICE Methods and Processes Across Health Technology Assessment Programmes: Why the Differences and What is the Impact?.. Applied health economics and health policy 15(4): 469-477	- Not about QES
DeJean, Deirdre, Giacomini, Mita, Simeonov, Dorina et al. (2016) Finding Qualitative Research Evidence for Health Technology Assessment. Qualitative health research 26(10): 1307-17	- Not about guidelines - No extractable methods
Dixon-Woods M.; Fitzpatrick R.; Roberts K. (2001) Including qualitative research in systematic reviews: Opportunities and problems. Journal of Evaluation in Clinical Practice 7(2): 125-133	- Not about guidelines
Eakin, Joan M and Mykhalovskiy, Eric (2003) Reframing the evaluation of qualitative health research: reflections on a	- No extractable methods

Study	Code [Reason]
review of appraisal guidelines in the health sciences. Journal of evaluation in clinical practice 9(2): 187-94	- Not about guidelines
Fadlallah, Racha, El-Jardali, Fadi, Nomier, Mohamed et al. (2019) Using narratives to impact health policy-making: a systematic review. Health Research Policy & Systems 17(1):	- Not about QES - Not about guidelines
Florez, Ivan D, Morgan, Rebecca L, Falavigna, Maicon et al. (2018) Development of rapid guidelines: 2. A qualitative study with WHO guideline developers. Health research policy and systems 16(1): 62	- Not about QES
Gargon, Elizabeth; Williamson, Paula R; Young, Bridget (2017) Improving core outcome set development: qualitative interviews with developers provided pointers to inform guidance. Journal of clinical epidemiology 86: 140-152	- Not about QES - No extractable methods
Grant, Sean, Hazlewood, Glen S, Peay, Holly L et al. (2018) Practical Considerations for Using Online Methods to Engage Patients in Guideline Development. The patient 11(2): 155-166	- Not about QES *[Not about using QES for guideline development.]*
Grummer-Strawn L.M. (2018) Development of evidence-based guidelines at the World Health Organization: A case-study of the ten steps to successful breastfeeding. Breastfeeding Medicine 13(7): a-3	- Conference abstract

Study	Code [Reason]
Harden, Angela, Thomas, James, Cargo, Margaret et al. (2018) Cochrane Qualitative and Implementation Methods Group guidance series-paper 5: methods for integrating qualitative and implementation evidence within intervention effectiveness reviews. Journal of clinical epidemiology 97: 70-78	- Not about guidelines
Huls S.P., Whichello C., van Exel N.J. et al. (2018) PATIENT PREFERENCES IN HEALTH TECHNOLOGY ASSESSMENT: A SYSTEMATIC REVIEW AND RESEARCH AGENDA. Value in Health 21(supplement3): 325	- Conference abstract
Kelson, Marcia, Akl, Elie A, Bastian, Hilda et al. (2012) Integrating values and consumer involvement in guidelines with the patient at the center: article 8 in Integrating and coordinating efforts in COPD guideline development. An official ATS/ERS workshop report. Proceedings of the American Thoracic Society 9(5): 262-8	- Not about QES - No extractable methods
Korhonen, Anne, Hakulinen-Viitanen, Tuovi, Jylhä, Virpi et al. (2013) Meta-synthesis and evidence-based health care - a method for systematic review. Scandinavian Journal of Caring Sciences 27(4): 1027-1034	- Not about guidelines
Krahn, Murray and Naglie, Gary (2008) The next step in guideline development: incorporating patient preferences. Jama 300(4): 436-438	- No extractable methods - Not about QES

Study	Code [Reason]
Langlois, Etienne V, Tunçalp, Özge, Norris, Susan L et al. (2018) Qualitative evidence to improve guidelines and health decision making. Bulletin of the World Health Organization 96(2): 79	- No extractable methods
Lewin, Simon, Glenton, Claire, Munthe-Kaas, Heather et al. (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): e1001895	- No extractable methods
Li Y., Yu J., Du L. et al. (2015) Exploration and practice of methods and processes of evidence-based rapid review on peer review of WHO EML application. Journal of Evidence-Based Medicine 8(4): 222-228	- Not about QES
Longworth, Louise, Sculpher, Mark J, Bojke, Laura et al. (2011) Bridging the gap between methods research and the needs of policy makers: a review of the research priorities of the National Institute for Health and Clinical Excellence. International journal of technology assessment in health care 27(2): 180-7	- Not about QES - No extractable methods
McPherson S., Rost F., Town J. et al. (2018) Epistemological flaws in NICE review methodology and its impact on recommendations for psychodynamic psychotherapies for complex and persistent depression. Psychoanalytic Psychotherapy 32(2): 102-121	- Not about QES - No extractable methods

Study	Code [Reason]
Murphy, E, Dingwall, R, Greatbatch, D et al. (1998) Qualitative research methods in health technology assessment: a review of the literature. Health technology assessment (Winchester, England) 2(16): iii-274	- Pre-2000 - Not about QES
Noyes, Jane, Booth, Andrew, Moore, Graham et al. (2019) Synthesising quantitative and qualitative evidence to inform guidelines on complex interventions: clarifying the purposes, designs and outlining some methods. BMJ global health 4(suppl1): e000893	- Not about QES
Opiyo, Newton, Shepperd, Sasha, Musila, Nyokabi et al. (2013) Comparison of alternative evidence summary and presentation formats in clinical guideline development: a mixed method study. PloS one 8(1): e55067	- Not about QES
Pope, Catherine; Van Royen, Paul; Baker, Richard (2002) Qualitative methods in research on healthcare quality. BMJ Quality & Safety 11(2): 148-152	- Not about QES
Richter Sundberg, Linda; Garvare, Rickard; Nystrom, Monica Elisabeth (2017) Reaching beyond the review of research evidence: a qualitative study of decision making during the development of clinical practice guidelines for disease prevention in healthcare. BMC health services research 17(1): 344	- Not about QES - No extractable methods

Study	Code [Reason]
Ring N, Ritchie K, Mandava L JR (2010) A guide to synthesising qualitative research for researchers undertaking health technology assessments and systematic reviews	- Duplicate reference
Roddis, Jennifer Karen, Liversedge, Hannah L, Ryder, Isobel et al. (2018) Incorporating the patient experience into clinical guidelines: recommendations for researchers and guideline developers. BMJ evidence-based medicine: bmjebm-2018	- No extractable methods
Rosedale, Mary, Malaspina, Dolores, Malamud, Daniel et al. (2012) Developing patient-centered treatment protocols in brain stimulation: a rationale for combining quantitative and qualitative approaches in persons with HIV. Journal of the American Psychiatric Nurses Association 18(3): 166-74	- Not about QES - Not about guidelines
Saunders, Hannele (2015) Translating knowledge into best practice care bundles: a pragmatic strategy for EBP implementation via moving postprocedural pain management nursing guidelines into clinical practice. Journal of clinical nursing 24(1314): 2035-51	- Not about QES - Not about guidelines - No extractable methods
Schünemann, Holger J; Fretheim, Atle; Oxman, Andrew D (2006) Improving the use of research evidence in guideline development: 10. Integrating values and consumer involvement. Health research policy and systems 4(1): 22	- Not about QES

Study	Code [Reason]
Staniszewska, Sophie, Boardman, Felicity, Gunn, Lee et al. (2014) The Warwick Patient Experiences Framework: patient-based evidence in clinical guidelines. International Journal for Quality in Health Care: journal of the International Society for Quality in Health Care 26(2): 151-7	- Not about QES
Tong A., Palmer S., Craig J.C. et al. (2016) A guide to reading and using systematic reviews of qualitative research. Nephrology Dialysis Transplantation 31(6): 897-903	- Not about guidelines
van Wesel, Floryt, Alisic, Eva, Boeije, Hennie et al. (2014) Using qualitative evidence to optimize child PTSD treatment guidelines. Psychological Trauma: Theory, Research, Practice, and Policy 6(5): 546-554	- Not about guidelines
Weich S, Fenton SH, Bhui K et al. Realist Evaluation of the Use of Patient Experience Data to Improve the Quality of Inpatient Mental Health Care (EURIPIDES) in England: study protocol. BMJ open 8(6): e021013	- No extractable methods - Not about QES
Wieringa, Sietse, Dreesens, Dunja, Forland, Frode et al. (2018) Different knowledge, different styles of reasoning: a challenge for guideline development. BMJ evidence-based medicine 23(3): 87-91	- Not about QES - No extractable methods
Zhang, Yuan, Coello, Pablo Alonso, Brozek, Jan et al. (2017) Using patient values and preferences to inform the importance	- Not about QES

Study	Code [Reason]
of health outcomes in practice guideline development following the GRADE approach. Health and quality of life outcomes 15(1): 52	- No extractable methods

APPENDIX E: SYSTEMATIC REVIEW OF THE METHODOLOGICAL LITERATURE FOR INTEGRATING QUALITATIVE EVIDENCE SYNTHESES INTO HEALTH GUIDELINE DEVELOPMENT – ACCEPTED VERSION

Carmona C, Baxter S, Carroll C. Systematic review of the methodological literature for integrating qualitative evidence syntheses into health guideline development. Res Syn Meth. 2021;1–15. https://doi.org/10.1002/jrsm.1483

Systematic review of the methodological literature for integrating qualitative evidence syntheses into health guideline development

Chris Carmona MSc

Susan Baxter PhD

Christopher Carroll PhD

School of Health and Related Research (ScHARR), University of Sheffield, UK

Correspondence

Chris Carmona

School of Health and Related Research

University of Sheffield

Sheffield

UK.

Email: ccarmona1@sheffield.ac.uk

Funding information

None

Abstract

Guidelines produced by local, national and international bodies underpin clinical practice and healthcare services worldwide. For guidelines to be based on the best available evidence, it is critical that syntheses of both qualitative and quantitative evidence are used to inform decision making. As methods for QES develop, they are increasingly able to

inform health guideline production. However, the process whereby this form of evidence is considered and incorporated tends to be unclear. This systematic review synthesised existing guidance concerning the use of QES in guideline development. Sources published in English that described or prescribed methods for incorporating QES into evidence-based health guidelines were eligible for inclusion. Seventeen relevant papers were identified. The literature indicates that there is a reasonable consensus about many stages of conducting a QES to inform guideline development. Areas needing further exploration include: the way that committees engage with QES; the usefulness of different QES methodologies; and understanding of how expert committees use evidence. Methods for producing QES for guideline committees tend to be similar to quantitative systematic review methods in terms of searching, quality appraisal, systematic management of data, and presentation of results. While this allows transparency and accountability, it could be argued that it is less 'true' to the principles of being led by the data, which are fundamental to most qualitative research. Understanding the process of using QES to produce guidelines is critical to determining their validity and applicability, and to ensure that healthcare provision is based on the best available evidence.

KEYWORDS

health guidelines, qualitative evidence synthesis, systematic review

1 INTRODUCTION

Guidelines produced by local, national and international bodies are used to underpin clinical practice and the delivery of healthcare services worldwide. The WHO for example lists 239 guidelines on its website[1], and the National Institute for Health and Care Excellence (NICE) in the United Kingdom currently lists 1,623 guidance products, including 354 guidelines[2]. Understanding the process of producing these guidelines is critical to determining their validity and applicability, and to ensure that healthcare provision is based on the best available evidence.

The use of findings from QES as evidence in the development of health guidelines is growing as the need for relevant and context-sensitive evidence increases[3]. This is commonly agreed to be because qualitative data can answer particular types of questions far better than quantitative data. Quantitative data are still key for questions of efficacy but are less able to

answer questions relating to understanding of patient preference, and other contextual outcomes such as feasibility and acceptability. These questions are best answered by qualitative studies[4]. QES might usefully answer questions that are key to guideline production, for example, how different groups of practitioners, people using services or stakeholders perceive the issue, what social and cultural beliefs, attitudes or practices might affect this issue or how different groups perceive the intervention or available options[5]. Increased recognition of the value of QES is also driven by the move towards greater patient-centredness in health systems, for example an emphasis on shared decision making, and the greater inclusion of patients and lay experts in guideline-producing committees[6]. It is also claimed that incorporating QES into guideline development can help to represent people who may otherwise be excluded from the process[7], and that QES can also potentially offer a valuable supplement to the experiences of patient representatives on guideline panels[8]. This does not mean that advocates of QES in guideline development are oblivious to the challenges of the approach. A number of concerns have been reported about using QES within international guideline-producing bodies. Firstly, authors acknowledge that guideline producers are principally systematic reviewers who may have no background or expertise in qualitative methods, and therefore need training to be able to produce high-quality QES[5,9,10]. Secondly, many qualitative researchers do not support the practice of synthesising qualitative research, and that for those that do there is no universally accepted way of doing this in health and social care[6] (though this position has changed somewhat since the publication of that paper). Standardisation of methods for producing QES is contrary to many qualitative approaches that are data led and iterative. There is a call for QES not to violate the underpinning epistemological foundations of the included studies[11]. Thirdly, qualitative research itself has been criticised by positivist authors as being context-dependent and specific, for including an insufficient number of informants, for being interpretative and, because it usually relies on small, purposive samples, for having a low degree of generalisation[12,13]. Conversely, this is regarded by qualitative researchers as one of the great strengths of qualitative research. It has also been argued that there are issues with the linking, mixing or merging of qualitative and quantitative evidence, and there is no ready-made toolkit for doing this[8]. The purpose of the current study however is to explore the methods of qualitative evidence synthesis in health guidelines, not to argue for or against their use.

A review of the use of qualitative data by NICE up to 2009, in addition noted a lack of consistency in terminology and method and even lack of agreement about what constituted a qualitative study across their different guideline-producing centres[10]. Other authors have highlighted a lack of clarity about the processes involved in how committees make decisions on the basis of qualitative evidence, and furthermore, how the strength of evidence relates to the strength of recommendation when QES is included[4,7]. For example, WHO guidelines have been criticised for making 'strong recommendations' despite there being only low or very low confidence in the underpinning evidence[4]. However, WHO argue that their guideline panels are expected to take into account broader evidence about acceptability, feasibility, and equity, in addition to evidence about effectiveness[4].

Methods for the synthesis of quantitative evidence are well established, and robust methods for meta-analysis and the pooling of quantitative data provide clearly interpretable information for decision making bodies. Interpretation of the available evidence is also supported by an established framework for determining its quality through use of the GRADE tool[14].

Alongside the ongoing concerns over their use and the readiness of guideline-producing bodies to integrate QES evidence into their processes, it is crucial to examine the methods that are being adopted or proposed both by experts in the field, and by guideline-producing bodies themselves. Recent growth and development in methods and standards for QES, and the development of tools to ensure standardisation both in QES (for example CERQual[15] and the work of the Cochrane QIMG[16]) and in guideline development (for example, the DECIDE collaboration Evidence-to-Decision frameworks[17,18]) has put qualitative evidence firmly on the agenda for evidence-based medicine. However, the most appropriate methods for using it during guideline development remain unclear. This study aimed to systematically review the methodological literature that addresses this topic, produce a synthesis of the state of the field, and explore where consensus and disagreement may exist.

2 METHODS

The review question was: what methods and processes are proposed in the methodological literature for incorporating QES into evidence-based health guideline development? As a methodological review the protocol was not eligible for PROSPERO registration, therefore it is provided as supplementary file 1.

2.1 Eligibility criteria

Papers, online sources or published manuals that described or prescribed methods for incorporating QES into evidence-based health guidelines were eligible for inclusion if they gave sufficient detail to allow extraction of different stages of the process and the methods used in those stages. Since the papers included were descriptive and not empirical studies, no study design restrictions were placed. Similarly, no country restrictions were put in place, although included papers needed to be published in English.

2.2 Information sources and search strategy

Health related databases were searched for papers published in English since 2000. The date was selected because neither qualitative evidence synthesis methodology, nor methods for guideline development were well established before that time. Database searches were conducted in MEDLINE (including MEDLINE in-process), EMBASE, CINAHL and PsycINFO from 2000 up until 15 August 2019. Supplementary searches were conducted in Google Scholar and results from the first six pages were added to the search results. Reference lists of papers were checked for further potential includes. A full search history can be found in supplementary file 2.

2.3 Data collection process

Titles and abstracts for all papers were screened, and those that appeared to meet the inclusion criteria, or those where it was uncertain whether the criteria were met or not were examined as full text. Articles marked for inclusion at this stage also had their reference lists checked to identify further papers. To maximise transparency of selection at the full text stage, a checklist was used to ensure that papers met the criteria for inclusion as described above (see supplementary file 3, for the completed checklist).

2.4 Data extraction

An *a priori* data extraction framework was set up in NVivo 12[19] to map the different stages of the reviewing process. The stages were chosen to represent the potential range of discrete tasks that are involved in a quantitative systematic review or QES, that is protocol/scope/review question; searching; study selection; data synthesis, critical

appraisal; quality appraisal; making recommendations, use of logic models/frameworks; integration with quantitative data/reviews; and reporting. Three additional 'umbrella' categories were also used to capture broader themes about the use of QES in guidelines - benefits of using QES in guidelines; challenges of using QES in guidelines; and QES methodologies that have been used in guideline development. Additional emergent themes were coded as they occurred during the data extraction process.

2.5 Risk of bias/quality appraisal

Formal risk of bias or quality assessment of included papers was not appropriate as they were methodological rather than empirical studies however, the design and any key weaknesses of included papers were noted during data extraction.

2.6 Method of synthesis of results

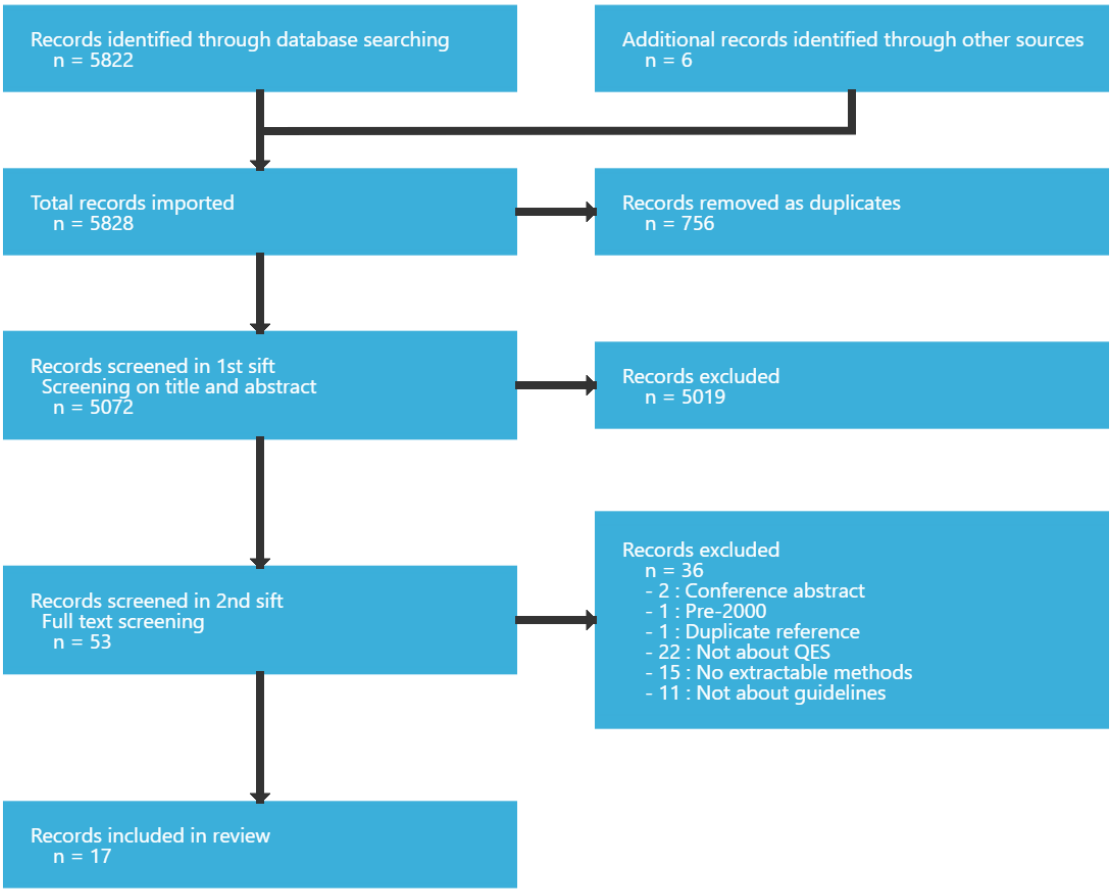
Data extracted was synthesised narratively, within the coding categories described above. Particular attention was paid to possible overlaps between different stages and the overarching 'umbrella' categories.

3 RESULTS

3.1 Study selection

Searching of databases yielded a total of 5,822 references. These were uploaded into EPPI reviewer 5 software[20] and de-duplicated. A total of 756 duplicate records were identified and removed. In total, 5,066 records were screened at title and abstract level. 5,019 records were excluded with 47 papers marked for full text examination. For a list of papers excluded at full text, along with reasons for their exclusion, see supplementary file 6. Eleven articles from data base searching and six additional references from reference list checking met the inclusion criteria. A total of 17 papers were included in the review. See figure 1 for a PRISMA diagram summarising the flow of papers.

Figure 1 PRISMA flowchart of included studies



3.2 Study characteristics

Full references of included papers are provided in supplementary file 5. Table 1 provides a brief summary of each paper.

Table 1 Brief details of included papers

Author	Type of paper	Summary
Booth 2016	Guidance	A report funded by the EU as part of a series on evaluating complex interventions ('INTEGRATE-HTA'). The guidance document sets out a framework to enable reviewers to choose between different QES

		methods depending on the question they are asking.
Carroll 2017	BMJ Analysis	The analysis focusses on the need for successful guidelines to reflect patient views and argues that qualitative evidence is a key way to do this. The paper is not primarily a detailed methodological paper but contains some extractable methodological detail.
Downe 2019	Research article	The first in a series of three papers that have been written by a group of methodologists working with the WHO on guidelines that integrated QES. The authors examine the use of QES in developing clinical and health systems guidelines.
Flemming 2019	Analysis	This paper presents an overview of the ways QES can be used to address complex interventions.
Glenton 2016	Manual/handbook	Chapter 15 of the WHO handbook for guideline development specifically about using evidence from qualitative research to develop WHO guidelines.
Glenton 2019	Research article	The third in a series of three papers describing the use of QES to inform the development of clinical and health systems guidelines by a team of methodologists who have worked with WHO. The WHO is increasingly using evidence derived from QES to provide information on acceptability and feasibility in its guidelines.
Gould 2010	Methodological report	Gould describes qualitative work done to support the production of two social care guidelines by the UK National Institute for Health and Care Excellence (NICE).

Hansen 2011	Methodological report	This article focuses on qualitative research synthesis in eliciting patients' perspectives as part of the growing drive to include patient views in policy and HTA.
Knaapen 2015	Toolkit (chapter)	Chapter 2 of a GIN toolkit on patient and public involvement in guidelines. It contains practical ideas about how to conduct a qualitative evidence synthesis as part of the guideline development process.
Kristensen 2007	Manual/handbook	The 2007 updated edition of the *Health Technology Assessment Handbook* that was issued by the Danish National Board of Health in 2001 as part of the fulfilment of the National Strategy for HTA. Contains some general detail about QES and also a specific chapter on assessment and syntheses [sic] of qualitative studies (s.4.2).
Lewin 2018	Commentary	Argues that the development of more 'robust' (transparent) methods and tools for QES has widened the opportunities for QES to be used to inform health guidelines (in the context of the WHO).
Lewin 2019	Research article	This is the second in a series of three papers written by methodologists working with the WHO that examines the use of QES in developing clinical and health system guidelines. It specifically discusses using qualitative findings as part of Evidence-to-Decision frameworks.
NICE Manual 2018	Manual	The process manual used by NICE to produce clinical guidelines. The NICE manual includes

		details of synthesis for all the types of evidence it uses, not just qualitative evidence.
Ring 2010	Guidance	Guidance from NHS Quality Improvement Scotland about the various methods of QES that could be used in HTA.
Ring 2011	Research article	The authors conducted a systematic search to identify QES and reflect on the methodological approach used.
Swedish Agency for Health Technology Assessment and Assessment of Social Services (SBU) 2016	Manual	Swedish agency for health technology assessment manual for evaluating and synthesising qualitative material.
Tan 2009	Evidence utilisation report	Describes the use of qualitative research as evidence in a national clinical guideline program (National Institute for Health and Clinical Excellence – NICE, UK).

3.3 Synthesis of results

3.3.1 Methods for review protocol development or scope generation

Seven of the 17 included papers discussed or referred to the need for a scoping process or a process of review protocol generation before the searching and literature identification phase of the development of a guideline. In terms of the scoping or review protocol development itself, the critical tasks for the scoping phase of a guideline are described as identifying the interventions, stakeholders and contexts relevant to the guideline questions. This can be time-consuming, and adequate time needs to be set aside for this part of the

process[7]; the decisions reached in these discussions directly inform the question and content of the review protocol.

Overall, there is broad agreement that a review protocol or scoping process should use expert input (for example a guideline committee, or service user organisation) to derive the review question and the criteria for that review (for example using a PICO or SPICE format). The protocol is therefore the first stage of the process unless prior searching is required for evidence to inform the scope or protocol (for example a QES). Scopes or protocols for QES may include a reflexivity statement.

The review protocol or scope should be made publicly available before the review commences in the same way as a quantitative systematic review would be registered[21].

3.3.2 Methods for identifying literature

Two thirds of the included papers contained information regarding the optimal methods of identifying evidence for a qualitative evidence synthesis. Most agreed that as part of the process there needed to be a systematic search of databases and pointed out that in many ways this was similar to quantitative database searching.

This focus on systematic searching represents a step-change from earlier in the 2000s when searching seems to have been less developed. However, some authors argue for systematic searching but also note that it may not be important to identify every available study, citing theoretical saturation as a possible endpoint[6,22]. QES conducted alongside a quantitative systematic review will be more likely to have more explicit inclusion criteria than a synthesis of qualitative studies that aims for theoretical saturation, where there might be a more iterative approach to searching and screening[6].

Purposive sampling is also suggested. It is described as an iterative process of searching and screening, with the process being complete when the reviewers achieve “theoretic saturation” [sic] or “conceptual robustness”[9]. It is frequently used as an adjunct to, or occasionally as a replacement for, exhaustive systematic database searching.

As a result of concerns over missing data, most authors recommend some kind of additional search method. Methods for additional searches that were mentioned include footnote and reference list checking, hand-searching key journals relating to the topic of the QES, forward citation searching (searching for relevant work by locating studies that cite earlier key

studies), and author searching (searching for all publications by the author of a relevant work).

In summary, while most authors advocate some kind of systematic searching process, the complexities of identifying qualitative literature have led to approaches that try to reduce the volume of literature found by comprehensive searches (for example by using filters or more specific search terms), while also trying to mitigate the potential loss of relevant papers by adding in supplementary search techniques such as citation searching.

3.3.3 Methods of study selection

Only three of the 17 papers gave any detail about considerations required in selecting qualitative studies. Two at some length[23,24], and the other[21] in a single paragraph. Fundamentally the authors agreed that the process for study selection of qualitative literature should mirror the process that would be expected in a quantitative systematic review, with multiple reviewers comparing the paper with pre-specified inclusion and exclusion criteria, with any disagreements resolved by discussion or by the use of a third reviewer. This was especially the case when QES was being conducted alongside a quantitative systematic review[24]. None of the other papers discussed methods of study selection.

Study selections should be transparently reported, for example through using a PRISMA diagram (Preferred Reporting Items for Systematic Reviews and Meta-analysis)[25] to show flow of studies through various stages of selection[23,24].

There is a possibility of retrieving a large number of studies, especially through systematic searching, and it is recommended that in these cases reviewers select a sample[23]. It is difficult to quantify what constitutes a 'large number' as it will, to a certain extent, depend on the emerging themes and concepts as well as on resources available and the timeframe of the review. Reviewers also need to be aware of introducing reviewer bias[23].

3.3.4 Methods of quality appraisal

The ten included studies that discuss critical appraisal are broadly in favour of assessing the methodological quality of the studies included in a QES. This position is likely to be strengthened by the introduction and widening use of the GRADE-CERQual tool[15] for

assessing confidence in findings from QES since CERQual relies on a methodological assessment (amongst other things) of the studies included in the review findings[9].

The latest edition of the NICE manual[22] states unequivocally that "Critical appraisal of qualitative evidence should be based on the criteria from the Critical Appraisal Skills Programme" (p. 106) and this is echoed by both in the WHO handbook[4] and the SBU methods manual[21]. More broadly, authors agree that studies should have some form of quality appraisal, preferably using one of the recognised appraisal systems for qualitative research[23]. The same advice is found in relation specifically to HTA[13]. Other authors, however, are more cautious and refer to lack of agreement about the value of critical appraisal of qualitative studies[6,24]. The GIN toolkit[9] succinctly summarises the issue – "The use of standardised 'checklist' approaches has been strongly critiqued by some commentators, questioning how quality criteria modelled on the principles of positivist science can be applied to non-positivist qualitative research" (p.33). However, in spite of this they provide a summary of the strengths and limitations of a range of checklists, including the CASP tool[26], the Cochrane manual (Chapter 20)[27], the cabinet office Quality in Qualitative Evaluation tool[28], and the Joanna Briggs Institute tool[29].

It is also noteworthy that some specific approaches to QES (for example framework, meta-narrative and thematic synthesis) all have their own approaches to critical appraisal, whereas in other approaches such as meta-ethnography or grounded theory, critical appraisal is less important[24]. The overall consensus is that some form of critical appraisal should be conducted that appraises the methodological conduct of the study. Methods for assessing the content and validity of data are discussed in section 3.3.8.

3.3.5 Methods of synthesis – General approaches

Much of the description of methods for the synthesis of findings from primary qualitative studies was presented as methods for specific QES methodologies. These specific approaches will be discussed after an outline of the generic methods referred to by other authors.

The Swedish HTA handbook for evaluation of qualitative studies[21] describes the evidence synthesis process as having four discrete stages. Firstly, papers are read to give an overview

of themes, then the papers are re-read and coded. No detail is given on the method of coding, but the manner in which it is described implies a process of emergent coding where codes are allowed to emerge from the included papers. As a second stage these 'first level themes' are "distilled to form the second level theme". This appears to be an aggregative coding process of drawing together similar codes. Thirdly, an interpretive coding phase is performed, described by the paper as follows - "Related second level themes are finally synthesised to an overall third level theme. Important patterns and associations among the second level themes are interpreted and problematised. The process is repeated until third level themes are set". The final stage in their synthesis process is "a general assessment of the scientific basis is made. Thereafter evidence graded results and conclusions are formulated", which appears to describe some assessment of review findings, like GRADE or GRADE-CERQual.

While there is agreement that findings should be 'added up' or compared and contrasted, the process of doing so inevitably masks their variability. This makes it easy to lose sight of the individuality of participants and their context that are the very heart of qualitative research. Two included papers highlighted how reviewers should strive to avoid this[9] and need to find the balance between splitting themes emerging from synthesis to the point that they are no longer useful, or lumping data together into themes that oversimplify or lose variation in the data[23].

This need for descriptive or interpretive themes is driven by the nature and purpose of the QES, which is in turn dependent on the nature of the guideline being developed[30]. The outputs from framework syntheses or thematic syntheses are often as simple as a list of themes identified across included studies, with little or no interpretation, that can be used to "detail the needs, values, perceptions, behaviours and experiences of stakeholders within the guideline"[23]. On a related note, it is suggested that QES used in guidelines tends to focus either on people's views about the interventions under scrutiny by the guideline, or it relates more widely to people's views and experiences of the condition underlying/addressed by the intervention(s) that the guideline is examining[7]. If interpretive findings are being produced, then there is a need for transparency on the part of the reviewers to ensure that the interpretations are plausible and to show how they were arrived at[23].

The use of Evidence-to-Decision frameworks (EtD) can be a driver for the style of the QES[23]:

> The main purpose of an EtD-orientated QES is to generate a series of findings from the included data, which are directly focussed on interventions addressed in the guideline, assessed for confidence and tailored towards acceptability, feasibility and equity, and the values that stakeholders attribute to the outcomes associated with the intervention. The findings are then added to the guideline EtD frameworks, prior to guideline panel consideration, (p.8)

Infographics and logic models can also be incorporated into EtD frameworks in cases where the synthesis is intended to be explanatory or theory building[7].

Overall, discussion about the general methods to be used for synthesis focuses on the continua between aggregative or integrative coding and more interpretive coding, and between the lumping and splitting of themes. This depends to some extent on the methodology used for the synthesis described in the next section.

3.3.6 Methods of synthesis - Specific methodologies

Several authors provide brief (or occasionally in-depth) descriptions of methods of synthesis that can be used. It is not the remit of this paper to reproduce general methodological detail about the various methods, but where authors have made comment on what makes a method suitable or unsuitable to produce a QES for a guideline development process, that has been included here.

There is a range of different QES methodologies available, some more developed than others. They predominantly reflect methods of primary qualitative research. The different methodologies sit broadly on a continuum between aggregative (or integrative) approaches that summarise themes and interpretive approaches that generate new interpretations of the data[30]. The Cochrane QIMG recommend that the method of synthesis should only be chosen after the evidence is known and caution against pre-specifying a methodology[31]. Epistemology is particularly important for some types of synthesis, with commentators arguing that the method of synthesis should be compatible with the epistemology of the included studies. Other methods may rely less on epistemology – for example best-fit framework synthesis, narrative synthesis and thematic synthesis. In health services research and technology assessment a more pragmatic approach is taken with it being common to integrate different types of study within a single synthesis[31].

Selection of an appropriate method is seen by authors as complex and dependent on many factors, especially the distinction between aggregative methods (where themes are integrated/aggregated) and interpretive methods (where the researchers try to add additional layers of interpretation over the data). The philosophical view of the researcher and purpose of the review can be driving factors[6], as can the need for "pragmatic and relatively rapid methods of qualitative evidence synthesis" that might fit better with guideline developers' timelines[8]. Framework, narrative, and thematic synthesis are highlighted as particularly useful for answering questions about the uptake of interventions and for integrating quantitative and qualitative findings. This may make them particularly useful for developers of clinical guidelines - NICE already use some form of thematic synthesis in some of their guidelines[8].

This identification of thematic synthesis methods as highly appropriate to guideline development is in line with the latest iteration of the NICE methods manual[22] which continues to identify thematic analysis as an appropriate methodology for analysing qualitative data. It advocates extracting 'first level themes' into evidence tables (Evidence tables are detailed summaries of the content of each study included in a review or synthesis. These are normally incorporated into an appendix of the review or synthesis.). These evidence tables are then used to generate 'second level themes' in the body of the synthesis. The manual also goes on to discuss (in passing) conceptual mapping, grounded theory, meta-ethnography and meta-synthesis, but notes that expertise in their use is needed (p.107).

Most of the studies that specify methodologies refer predominantly to the same pool of synthesis methods: narrative synthesis[12], meta-synthesis[9,13], 'imported concepts'[13], meta-ethnography[9,13], meta-study[13], qualitative meta-summary[9,13] and framework analysis[9]. The older Danish HTA manual[12] mentions only meta-ethnography and narrative synthesis. This is likely because limited QES methodologies were available at that time. In a 2011 survey of 107 different QES, reviews using critical interpretive synthesis, meta-interpretation, qualitative cross-case analysis and grounded theory synthesis were found infrequently, and therefore their usefulness as methods of QES for HTA is unknown[6].

By far the most comprehensive and well-developed guidance for selecting an appropriate QES method for HTAs is a report for the INTEGRATE-HTA project[31]. This project develops

various criteria for QES and matches them to 19 different methodologies. Reviewers can select an appropriate method by aligning their needs with the various methods for conducting QES as outlined in the guidance's comprehensive tables that clarify a diverse range of considerations for each method. The project was directed specifically at HTA methods, but there seems no reason to suppose they would not be equally applicable to broader health guidelines.

Overall, thematic synthesis is the most frequently mentioned form of QES in guidelines and seems to be the most commonly used, with meta-ethnography and framework or best-fit framework synthesis as alternatives. This is primarily because these are the methods that are easier to use, and other methods have not been well tested, so they may be useful or not[6]. While a broad range of QES methodologies can be useful, the art is in selecting the appropriate methodology for the research question and research context[31]. The method of synthesis should only be chosen after the evidence is known[30].

3.3.7 Reporting standards

Reporting standards for QES were not discussed at great length in any of the included papers, possibly because different organisations have well established reporting standards internally. However, "…generic qualitative evidence synthesis reporting guidelines exist, others are being developed for particular methods, and standards are evolving to establish the level of confidence users can ascribe to the findings of such syntheses."[8]

There is a trend towards systematic review-like transparency in QES[23,30]. Historically, transparency has not been handled well by people reporting QES, but work has been undertaken to develop reporting standards for QES, such as the ENTREQ tool[32] and the eMERGe tool for meta-ethnography[33]. Newer tools have also emerged, notably RAMESES for realist synthesis[34].

A useful minimum reporting standard has been used in work with the WHO[23]. The standard closely matches the reporting standards for quantitative systematic reviews (Cochrane reviews particularly) and suggests the characteristics and critical appraisal of each study should be presented in some detail, accompanied by a summary of themes (summary of qualitative findings) along with the confidence in those review findings and reasons for any downgrading. It also suggests providing a list of excluded studies, along with reasons for exclusion.

The guideline handbooks also briefly recommend approaches to reporting, with the WHO handbook[5] recommending the use of a summary of qualitative findings table that includes CERQual assessment (if there is one). SBU[21] also recommends the use of tabulation and the use of illustrative quotes where possible. The NICE manual[22] is more prescriptive and requires researchers to provide extensive evidence tables for all included studies containing the key information about the study. When CERQual is not being used, the NICE manual requires the production of evidence statements that summarise the evidence, its context and quality, and the consistency of key findings and themes across studies (meta-themes).

3.3.8 Moving from evidence to recommendations

Papers discussed various aspects of the process of evidence-based recommendation-making that fall generally into four categories - certainty in findings (including CERQual); frameworks (including Evidence-to-Decision frameworks); committees; and making recommendations from the evidence.

Certainty in findings from QES

Many guideline development agencies, including WHO and NICE require information on the confidence of findings that are used to underpin recommendations[7,22]. Since its publication in 2015, use of GRADE-CERQual has become the most common tool used as a summary measure when evaluating qualitative evidence for guideline development. A series of papers based on a WHO guideline were written by members of the original team who authored and devised the CERQual system for assessing the certainty in findings of qualitative evidence, and in the WHO papers they recommend the use of CERQual in guideline development[4,7,23]. The earliest mention of CERQual in the included papers is in the WHO guideline handbook qualitative chapter[5], which contains a brief description of the components of CERQual as a tool to measure the level of confidence, in each of the findings of the QES. It also notes its similarity to GRADE for quantitative studies.

NICE recommend the use of CERQual somewhat more robustly. The manual notes that unless the qualitative evidence is very sparse or disparate (in which case a narrative approach is appropriate), the results of QES should be presented as summaries (‘at outcome

level') and should be assessed with GRADE-CERQual. They present 'evidence statements' (narrative summaries) as a less preferred alternative[22].

Frameworks

The three papers in the recent WHO series[4,7,23] discuss at some length the use of Evidence-to-Decision frameworks (EtD), an approach developed by the GRADE working group to increase transparency in moving from evidence and contextual considerations to implementable recommendations[17,18]. These EtD frameworks take the form of tables that draw together the key information necessary for guideline committees to make recommendations, including the PICO for the research question, summaries of the evidence, details of equality issues, feasibility issues, implementation consideration etc. They contribute to the overall transparency of the movement from evidence through discussion by a guideline committee or similar into recommendations but are not specific to QES. The evidence from QES can be added to the evidence section of the EtD framework alongside any quantitative evidence, along with its CERQual assessment[6], Qualitative evidence does not always fit well within the" summary-based and compartmentalised structure" of the EtD framework. There are also implementation issues related to clinical guidelines, and it is possible that evidence from QES that does not make it into the evidence section of the EtD framework can often be rewritten a little and turned into an implementation consideration[4].

The only paper outside of the three WHO papers that discussed EtD frameworks stated that "[a] QES can be conducted separately or can be integrated with some form of quantitative synthesis. Within a guideline development process, findings from a QES will often be integrated with evidence of effectiveness in an evidence-to-decision (EtD) framework, used to formulate recommendations"[30]. In another paper, a broader discussion about frameworks generally, the usefulness of frameworks in organising data for a QES, and also for identifying gaps in qualitative data is highlighted[4].

Committees

There was little discussion of the role of any kind of guideline or oversight committee in interpreting the evidence generated by QES and using it to develop recommendations, beyond the discussion reported above in relation to EtD frameworks. One example described is a process undertaken in the production of a social care focussed guideline (on dementia) where the evidence was searched for and reviewed by an academic review team,

but the “weighting and synthesis” of evidence was done jointly with a guideline committee that included patients and carers[35].

During the process of guideline production in committees, members may need to be reminded of relevant qualitative evidence, and this ‘champion’ role might more easily be taken up by the producer of the synthesis, the methodologist or patient representatives[9].

Making recommendations from the evidence

The process of making recommendations using the results of QES was not discussed in-depth in any of the papers, with those that mention it mostly reporting that it is difficult to capture by simple steps and rules[9]. Normally, committees (in whatever form they take) make recommendations based on one or more systematic reviews, including any QES, alongside any other information that the committee consider to constitute ‘evidence’ (for example EtD frameworks). However, this is not always the case and points out that sometimes confidence in QES or other types of evidence-based on published studies may be overridden, for example by human rights considerations or other overarching principles or normative values[4].

So, although the evidence is primary, it is not the only consideration for guideline committees, and that is true of both quantitative systematic reviews and QES. The amount that any kind of evidence drives a decision about a particular recommendation should depend on the question being considered, and the judgments made should be supported by clear and transparent justifications[6]. There is little contained in the included papers to explain or clarify the process of using QES to inform recommendations.

4 DISCUSSION

Overall, the literature relating to the use of QES within the context of guideline development seems to mirror large parts of the more general literature on QES, and this is of little surprise since the key people driving the development of QES methods in health and social care are also often the same people who are driving the agenda for using QES in guideline development processes.

In a world where evidence-based healthcare is dominated by the systematic review of randomised controlled trials (RCTs) by organisations such as NICE, Cochrane, WHO etc., a model of QES that matches their already existing methods of standardisation is likely to be more acceptable to them (and to fit better with their existing methods of interrogating

evidence) and standardisation of QES methods may be, in part, a strategic move by their advocates to make the methods more acceptable to organisations that have traditionally been sceptical of qualitative research. The Cochrane Qualitative and Implementation Methods Group (QIMG) have been instrumental in increasing the standardisation of QES, most notably by supporting the development of GRADE – CERQual for assessing the level of confidence in summary qualitative findings in a way that clearly (and purposefully) matches the process of using the GRADE tool on quantitative pooled outcomes, and also by supporting methods of QES that can be integrated into or presented alongside Cochrane systematic reviews.

There are some sections of the review process where much of the literature was silent, for example only three papers discussed study selection. It is unclear whether this was because they regarded it as less important, but none advocated a more traditional, emergent qualitative approach.

Overall, in this included literature, there is a level of excitement about the possibilities of using syntheses of qualitative evidence alongside traditional quantitative evidence, leading authors to cautiously proclaim a 'new era' for qualitative research, supported by recent developments in QES methodologies, such as standardised methods for synthesis and for assessing the confidence that can be placed in the findings[36].

In spite of this enthusiasm, there is still not universal agreement that qualitative evidence can be synthesised in a way that is meaningful or useful, or that standardisation of methods of QES to make them more acceptable to the evidence-based medicine movement is the best way to synthesise qualitative evidence. None of the included studies commented on the usefulness of more traditional reviews of qualitative evidence in developing guidelines. In large part, disagreements stem from the fact that primary qualitative evidence makes no claim to be generalisable, yet for a QES to be useful to a guideline-producing committee, the committee needs to be able to argue that the evidence speaks to common experience. Although these arguments seem to be broadly ignored by researchers producing QES currently, the early days of QES were dogged by these arguments[37] since it "has emerged from the confluence of conventional systematic review methods with methods for primary qualitative research. With such a mixed heritage, and the juxtaposition of quite different epistemological positions, it is inevitable that the resultant tensions have generated considerable creative energy and significant methodological frictions."[38]

Once past the epistemological arguments about the philosophical feasibility of QES, the practical methods of performing a QES seem to have converged, in terms of their applicability to guidelines at least, over the past decade. There seems to be broad agreement over most stages of producing a QES to inform a guideline, even if the fine detail is not always consistent. Table 2 highlights similarities and areas of agreement in the reviewed literature as well as some differences and gaps where further investigation could be fruitful.

Table 2 Areas of agreement and opportunities for further development

Areas of agreement	Gaps/assumptions	Development opportunities
Protocol development		
• Review protocols are important, using SPICE, PerSPEecTiF or SPIDER[38] rather than PICO formats • Beneficial to involve lay-people and experts in protocol development, but this is resource intensive and time-consuming • Different frameworks for formulating research questions/protocols are developing, for example recent work on the PerSPEcTiF framework.[41]	• The role of tertiary reviews ('reviews of reviews') is established in the quantitative literature (for example, the Cochrane manual[27], Chapter 22), and is often used for scoping reviews or mapping reviews to provide an overview of the field. There is no discussion of tertiary QES. • There is an unspoken assumption that a single QES would underpin an entire guideline.	• What is the value (if any) of syntheses of existing QES? • How useful is a single generic QES for a guideline compared to specific QES for different elements of the guideline.
Identifying literature		

• structured searching using validated qualitative filters • some kind of supplementary searching is also common - reference list searching, citation searching, asking experts, or trawling grey literature • some support for introducing a concept of theoretical saturation to prevent searching becoming too onerous	• Relevant data may be included in studies that are not directly relevant to the research question at hand	• What is the optimum balance between inclusive searching and specific searching and sifting to identify relevant themes for QES?
Study selection		
• process mirrors quantitative –studies are matched against inclusion and exclusion criteria in the review protocol, usually in a two-stage process, firstly based on title and abstract, and then for papers that are not obviously excludable, at full text.	• Assumes that standardised pre-specified protocol methods are superior to iterative, emergent qualitative methods.	• Can an iterative approach to study selection be transparent enough to meet the transparency requirements for health guidelines?
Quality appraisal		

• researchers undertaking QES for the purposes of guideline development are more often in favour of transparent and systematic methods than researchers with a more pluralist approach to QES • agreement about the importance of using some kind of transparent process for quality appraisal is greater within the QES for guidelines literature reviewed for this study than among the general QES literature • CASP most commonly used tool though over 100 tools in circulation	• little agreement about the best way to measure the quality of a qualitative study because of methodological variation and differing views about what a 'good' quality study looks like	• CAMELOT critical appraisal tool in development • Should studies of very low quality be excluded from any analysis or can they contribute to the overall analysis?
Synthesising findings		
• the most common (and probably the most accessible) form of synthesis is a thematic analysis that goes through a process of aggregative coding,	• It is uncertain whether some methods are more appropriate than others for particular types of question • Most methods of QES have not been used often in guideline	• Continuing evaluations of different QES methods used to underpin health guidelines

sometimes followed by interpretive coding	processes, and therefore it is unclear whether they are useful	
Reporting		
• lack of detailed discussion of reporting standards • the ENTREQ tool is commonly used as is some adaptation of the PRISMA standards for quantitative reviews. The Cochrane EPOC group have recently released a template for QES[42]	• Different organisations have their own in-house reporting standards. It is uncertain to what extent these overlap	• How well do different reporting standards from major QES producers differ, and how can the differences be resolved?
Recommendations		
• GRADE and GRADE-CERQual have been a valuable addition to the decision-making process, but there is often no obvious link between the review findings (and any other information considered by the guideline-producing bodies as evidence) and recommendations made by the committee (and	• It remains unclear how guideline committees/panels move from QES findings to making recommendations • Lack of understanding about the processes by which committees use evidence to generate recommendations – no clear insight into committee use of different types of	• How do guideline-producing committees engage with different types of evidence (alongside their own beliefs, knowledge and experience), and how do they use that evidence to form recommendations?

the relative strength of the recommendation) • only WHO seem committed to using EtD frameworks	evidence (quantitative, qualitative) to approach different types of questions (effectiveness, feasibility, etc.)	

4.1 Gaps highlighted by the review findings

This review has highlighted some areas where there appears to be enough consistency between different approaches to give a reasonable level of confidence in them. In some cases, there seems to be better agreement between authors writing about QES methods for guideline production than for QES more generally, for example in terms of the level of agreement around searching and around the need for critical appraisal of qualitative studies, both of which are contentious in the broader field of QES[39].

There are also areas where there is less clarity, some of which are likely to be quite specific to QES in the context of guideline production and therefore could probably not be resolved by wider searching of the literature.

As more and more QES are published in health and medicine, it becomes more likely that reviewers will find existing QES that wholly or partially answer their research questions. There is no discussion in the literature to explore how these may be used. Parallels in quantitative systematic reviewing include updating and using pre-existing reviews as evidence for committees; other reviewers use the inclusion lists from systematic reviews as a check that they have identified the relevant literature. In areas where several very similar systematic reviews exist a 'review of reviews' or tertiary review can be conducted. None of those things is reported in the papers included here, but similar methods might be possible for QES.

Much of the literature included in this review contains the unspoken assumption that one guideline will require one QES. Only one included paper moots the possibility of multiple QES for one guideline[7], however it is easy to imagine a guideline that contained questions

that could be informed by several QES. There is no discussion in the included papers of how this might work in practice.

In terms of producing QES that are useful to guideline committees as part of the evidence base they consider, is a standardised methodology best? Or is methodological pluralism more useful where the methodology can be a more pragmatic choice and take into account the time, resource and outputs that are wanted? There are some 30 methods for conducting a QES[31], and at least a further ten methods are in development[40]. Few of those methods have been used frequently in producing QES for guideline development and so their utility is uncertain.

A related gap in this review is an understanding of the most useful way to provide guideline committees with the outputs of various reviews, both qualitative and quantitative alongside other kinds of evidence. Traditionally, guideline producers have prioritised reviews of RCT evidence to provide evidence of clinical efficacy, but the hierarchy of evidence for other types of outcomes is less established. As QES become more standardised in their methods and transparency, can they become the principal evidence for certain types of guideline question? One series of papers discusses the use of EtD frameworks in some depth in one of the papers[7], but these have not been the subject of robust evaluation, and it is unknown how useful they are to committees.

Finally, there is little research, and none in this review, that explores how committees move from QES findings (or indeed quantitative systematic review findings) and the other information they are given as evidence – systematic reviews, expert testimony, real-world data etc, in the context of their own expertise and experience - to making decisions that produce guideline recommendations, and this is a fundamental question for guideline producers.

4.2 Limitations

The content of this review was specifically limited to papers describing methods for using QES in guideline production. As a result, it does not cover the large, and growing, corpus of literature dealing with the topic of QES generally. There are good, published overviews of the development of QES, notably through leadership from the Cochrane QIMG[16].

Fortunately, several of the key authors writing about the use of QES in guideline

development are also key authors in the field of QES in health more generally, so to a large extent the literatures inform and reflect each other.

For the purposes of time and resource, this review did not consider the implications of the growth in methods for integrating qualitative and quantitative data to produce mixed-methods reviews. As technologies for producing standardised QES and standardised systematic reviews develop, researchers are becoming interested in integrating qualitative and quantitative data in the aspiration that the whole may be greater than the sum of the parts.

5 CONCLUSIONS

The use of qualitative evidence syntheses to inform the production of health guidelines is growing as the methods for producing them become more clearly defined and more standardised. Methods for producing QES for guideline committees tend to be similar to quantitative systematic review methods in terms of searching, appraisal of evidence, systematic management of data and presentation of results. While this allows greater transparency and greater accountability, it could be argued that it is less 'true' to the principles of being led by the data, which are fundamental to most qualitative research.

Recent developments in QES mean that there is broad agreement about how QES can be produced to help inform guidelines, but further research is needed to establish whether guideline-producing committees find QES useful to their deliberations, whether they could be done or presented differently to make them more useful and, perhaps most importantly, how committees use QES to inform their decision making alongside traditional systematic reviews of effectiveness.

COMPETING INTERESTS

Chris Carmona is a part-time employee of the UK National Institute for Health and Care Excellence.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the supplementary material of this article.

HIGHLIGHTS

- There is a reasonable level of consensus about many of the stages of conducting a qualitative evidence synthesis to inform guideline development.
- However, the way that committees engage with QES, the usefulness of different QES methodologies for informing health guidelines and understanding of the way that expert committees use different types of evidence when developing guideline content requires further clarification.
- Understanding the process of using QES to produce these guidelines is critical to determining their validity and applicability, and to ensure that healthcare provision is based on the best available evidence.

SUPPLEMENTARY MATERIAL

Supplementary files 1 – 6 are included below within this appendix.

REFERENCES

1. World Health Organization. WHO guidelines approved by the Guidelines Review Committee. https://www.who.int/publications/guidelines/atoz/en/. Accessed 19/07/2020.
2. National Institute for Health and Care Excellence (NICE). Guidance and advice list. https://www.nice.org.uk/guidance/published?type=apg,csg,cg,cov,mpg,ph,sg,sc,dg,hst,ipg,mtg,qs,ta. Accessed 19/07/2020.
3. Langlois EV, Tunçalp Ö, Norris SL, Askew I, Ghaffar A. Qualitative evidence to improve guidelines and health decision making. *Bull World Health Organ.* 2018;96(2):79-79A.
4. Glenton C, Lewin S, Lawrie TA, et al. Qualitative Evidence Synthesis (QES) for Guidelines: Paper 3 - Using qualitative evidence syntheses to develop implementation considerations and inform implementation processes. *Health Res Policy Syst.* 2019;17(1):74.

5. Glenton C, Lewin S, Norris SL. Using evidence from qualitative research to develop WHO guidelines (Chapter 15). In. *World Health Organization Handbook for Guideline Development.* 2nd ed. Geneva: WHO; 2016.
6. Ring N, Jepson R, Ritchie K. Methods of synthesizing qualitative research studies for health technology assessment. *Int J Technol Assess Health Care.* 2011;27(4):384-390.
7. Lewin S, Glenton C, Lawrie TA, et al. 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.* 2019;17(1):75.
8. Carroll C. Qualitative evidence synthesis to improve implementation of clinical guidelines. *BMJ (Clinical research ed).* 2017;356:j80.
9. Knaapen L, Colvin CJ, Cowl J, van der Weijden T. How to include qualitative research on patient views in guidelines. *GIN public toolkit.* 2015.
10. Tan TPY, Stokes T, Shaw EJ. Use of qualitative research as evidence in the clinical guideline program of the National Institute for Health and Clinical Excellence. *International journal of evidence-based healthcare.* 2009;7:169-172.
11. Sandelowski M, Barroso J. *Handbook for synthesizing qualitative research.* springer publishing company; 2006.
12. Kristensen FB, Sigmund H, eds. *Health Technology Assessment Handbook* Copenhagen: Danish Centre for Health Technology Assessment, National Board of Health 2007.
13. Hansen HP, Draborg E, Kristensen FB. Exploring qualitative research synthesis: the role of patients' perspectives in health policy design and decision making. *Patient.* 2011;4(3):143-152.
14. Schünemann HJ, Brożek J, Guyatt GH, Oxman A, eds. *GRADE handbook for grading quality of evidence and strength of recommendations. [Updated October 2013].* The GRADE Working Group; 2013.
15. GRADE-CERQual: Confidence in the Evidence from Reviews of Qualitative Research. https://www.cerqual.org/). Accessed 14/09/2019.
16. Cochrane Qualitative and Implementation Methods Group. https://methods.cochrane.org/qi/welcome. Accessed 14/10/2019.

17. Alonso-Coello P, Schünemann HJ, Moberg J, et al. GRADE Evidence-to-Decision (EtD) frameworks: a systematic and transparent approach to making well-informed healthcare choices. 1: Introduction. *BMJ.* 2016;353:i2016.
18. Alonso-Coello P, Oxman AD, Moberg J, et al. GRADE Evidence-to-Decision (EtD) frameworks: a systematic and transparent approach to making well-informed healthcare choices. 2: Clinical practice guidelines. *BMJ.* 2016;353:i2089.
19. *NVivo qualitative data analysis software* [computer program]. Version 122018.
20. *EPPI Reviewer* [computer program]. Version 52018.
21. Swedish Agency for Health Technology Assessment and Assessment of Social Services (SBU). *Evaluation and synthesis of studies using qualitative methods of analysis.* Stockholm: Swedish Agency for Health Technology Assessment and Assessment of Social Services (SBU); 2016.
22. National Institute for Health and Care Excellence (NICE). *Developing NICE guidelines: the manual.* Manchester: NICE; 2018.
23. Downe S, Finlayson KW, Lawrie TA, et al. 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.* 2019;17(1):76.
24. Ring N, Ritchie K, Mandava L, Jepson R. *A guide to synthesising qualitative research for researchers undertaking health technology assessments and systematic reviews.* Glasgow: Quality Improvement Scotland (NHS QIS). 2010.
25. Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA). www.prisma-statement.org. Accessed 14/09/2019.
26. Programme CAS. CASP Qualitative Checklist. https://casp-uk.net/casp-tools-checklists/ Published 2018. Accessed 14/09/2019, 2019.
27. Higgins J, Thomas J, Chandler J, et al. *Cochrane Handbook for Systematic Reviews of Interventions.* 6 ed: Cochrane; 2019.
28. Spencer L, Ritchie J, Lewis J, Dillon L. Quality in Qualitive Evaluation: A framework for assessing research evidence. 2003. https://webarchive.nationalarchives.gov.uk/20140619055846/http:/www.civilservice.gov.uk/wp-content/uploads/2011/09/a_quality_framework_tcm6-38740.pdf. Accessed 14/09/2019.

29. Joanna Briggs Institute. JBI Critical Appraisal Checklist for Qualitative Research. 2017. https://joannabriggs.org/ebp/critical_appraisal_tools. Accessed 14/09/2019.
30. Flemming K, Booth A, Garside R, Tunçalp O, Noyes J. Qualitative evidence synthesis for complex interventions and guideline development: clarification of the purpose, designs and relevant methods. *BMJ Global Health.* 2019;4.
31. Booth A, Noyes J, Flemming K, et al. Guidance on choosing qualitative evidence synthesis methods for use in health technology assessments of complex interventions. [Online]. In: Integrate - HTA; 2016.
32. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Medical Research Methodology.* 2012;12(1):1-8.
33. France EF, Cunningham M, Ring N, et al. Improving reporting of meta-ethnography: the eMERGe reporting guidance. *BMC Medical Research Methodology.* 2019;19(1):1-13.
34. The RAMESES Project. https://www.ramesesproject.org/. Accessed 14/09/2019.
35. Gould N. Integrating qualitative evidence in practice guideline development: Meeting the challenge of evidence-based practice for social work. *Qualitative Social Work: Research and Practice.* 2010;9:93-109.
36. Lewin S, Glenton C. 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.* 2018;17.
37. Barnett-Page E, Thomas J. Methods for the synthesis of qualitative research: a critical review. *BMC Medical Research Methodology.* 2009;9(1):1-11.
38. Booth A. *Acknowledging a Dual Heritage for Qualitative Evidence Synthesis: Harnessing the Qualitative Research and Systematic Review Research Traditions.* , University of Sheffield.; 2013.
39. Carroll C, 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(2):149-154.
40. 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.

41. Booth A, Noyes J, Flemming K, Moore G, Tunçalp Ö, Shakibazadeh E. Formulating questions to explore complex interventions within qualitative evidence synthesis. *BMJ Glob Health.* 2019;4(Suppl 1):e001107.

42. Glenton C, Bohren M, Downe S, Paulsen E, Lewin S. EPOC Qualitative Evidence Synthesis: Protocol and review template. 2019.

Supplementary material

Contents

Supplementary file 1: review protocol

Supplementary file 2: systematic search history

Supplementary file 3: study inclusion checklist

Supplementary file 4: Summary of included papers

Supplementary file 5: included papers – full references

Supplementary file 6: papers excluded at full text

Supplementary file 1: review protocol

Field	Content
Review title	A systematic review of the methodological literature for integrating qualitative evidence syntheses into guideline development
Review question	What methods and processes are proposed in the methodological literature for incorporating QES into evidence-based health guideline development?

Objective	To capture the published and key grey literature that describe methods for incorporating QES into evidence-based health guidelines. This structured methodological review will capture primary data from coding of relevant papers using a template approach (described below) focussing on key methodological elements of the process: • Question design • Searching • Sifting • Coding/data extraction • Data synthesis • Use of logic models/frameworks • Integration with quantitative data/reviews • Decision making processes
Searches	The following databases will be searched: • Embase • MEDLINE • CINAHL • PsycINFO See appendix B for full search strategies for all databases. Search coverage will be checked using sentinel papers. Other searches: • Google – first six pages

	• Reference list screening of all included papers • Citation searching of all included papers Grey Literature will be included if cited on a reference list at screening or identified through the Google search. If less than ten papers are identified by the methods above, supplemental searches will be conducted as follows: • Hand-searching of key journals • Opportunistic methods: o Websites of research groups involved with QES in health or guideline development o JISCmail QES list o Expert consultation The searches will be re-run 6 months before final submission and further papers retrieved for inclusion. The full search strategies for MEDLINE database and full detail of additional and opportunistic searches will be published in the final review.
Condition or domain being studied	Qualitative evidence synthesis in evidence-based health guidelines.
Target	**Inclusion**: Papers that describe or propose methods and processes for incorporating QES into different

	kinds of evidence-based health (clinical or public health) or social care guideline. **Exclusion**: • Papers that do not describe a clear, extractable methodology or process. • Qualitative evidence syntheses themselves.
Content	Formal methods for qualitative evidence synthesis as part of the process of producing an evidence-based guideline. For the purposes of this review a qualitative evidence synthesis is defined as: a process of systematically and transparently combining evidence from individual qualitative studies to create new understanding by comparing and analysing concepts and findings from different sources of evidence with a focus on the same topic of interest An evidence-based guideline is defined as a series of statements or recommendations intended to optimise care or to inform decisions made by providers and service users. At a minimum: • the process should include a comprehensive and systematic assessment of the best available evidence • the evidence should be subject to a transparent decision-making process • the process should produce statements or recommendations, based on

	interpretation of the evidence (including the strength and quality of the evidence) that intend to improve care or support health decision making.
Comparator/Reference standard/Confounding factors	Not applicable
Types of study to be included	Papers, online sources or monographs that describe or prescribe methods for incorporating QES into evidence-based health guidelines. Sources must give sufficient detail to allow extraction of different stages of the process and the methods used in those stages.
Other exclusion criteria	• Non-English language • Books • Theses
Context	Any
Primary outcomes (critical outcomes)	Description of the methods used to perform the QES and integrate it with other data to inform the decisions made, specifically: • Question design • Searching • Sifting • Coding/data extraction • Data synthesis

	• Use of logic models/frameworks • Integration with quantitative data/reviews • Decision making processes
Secondary outcomes (important outcomes)	• Description of obstacles and challenges • Optimal methods • Adverse outcomes • Other emergent themes
Data extraction (selection and coding)	All references identified by the searches and from other sources will be uploaded into EPPI reviewer 5 or EndNote and de-duplicated. The full text of potentially eligible papers will be retrieved and will be assessed in line with the criteria outlined above. Papers suitable for inclusion will be uploaded into NVivo 12 and coded as described below
Risk of bias (RoB) (quality) assessment	Formal RoB assessment is not applicable since these are not primary studies, however the author may comment on the design and key weaknesses of included papers.
Strategy for data synthesis	Data will be extracted by open coding in NVivo 12, using a template codeset that is designed to capture the different methodological stages of guideline development and the processes of integrating QES with other data to inform healthcare decision making in guidelines. Papers will also be coded for emergent themes to allow for unanticipated relevant detail to be

	captured. Further data analysis may take place outside of NVivo using Microsoft Office packages.
Analysis of sub-groups	If there is sufficient consistency of named approaches, codes will be grouped by approach to explore whether different approaches offer different benefits and difficulties.
	☐ Diagnostic ☐ Prognostic ☒ Qualitative ☐ Epidemiologic ☐ Service Delivery ☐ Other (please specify)
Funding sources/sponsor	
Conflicts of interest	None
Dissemination plans	
Keywords	Qualitative evidence synthesis, Clinical guidelines, Public health guidelines, Social care guidelines
Details of existing review of same topic by same authors	None

Supplementary file 2: systematic search history

Topic/question details:

What methods and processes are proposed in the methodological literature for incorporating QES into evidence-based health guideline development?

Databases	Date searched	Version/files	No. retrieved
Embase (Ovid)	15/8/19	Embase <1996 to 2019 Week 32>	3727
MEDLINE (Ovid)	15/8/19	Ovid MEDLINE(R) <1996 to August 13, 2019>	719
MEDLINE In-Process (Ovid)	15/8/19	Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <1946 to August 13, 2019>	63
CINAHL(EBSCO)	15/8/19	CINAHL 1998 to 2019	884
PsycINFO (Ovid)	15/8/19	PsycINFO <2002 to August Week 1 2019>	410

Search strategies

Database: MEDLINE
Results: 719 Database: Ovid MEDLINE(R) <1996 to August 13, 2019> Search Strategy: --- 1 (("semi-structured" or semistructured or unstructured or informal or "in-depth" or indepth or "face-to-face" or structured or guide) adj2 (interview* or discussion* or questionnaire*)).tw,kw. (86169) 2 (focus group* or qualitative or ethnograph* or fieldwork or thematic or phenomenological or "grounded theory" or "field work" or "key informant").tw,kw. (192379) 3 interviews as topic/ or focus groups/ or narration/ or qualitative research/ (113878) 4 1 or 2 or 3 (283347)

5 ("meta-synthesis" or "Meta-synthesis" or "framework synthesis").tw,kw. (687)

6 exp Review Literature as Topic/ (12145)

7 (review* or overview*).ti. (304571)

8 (systematic* adj5 (review* or overview*)).tw. (125996)

9 (qualitative* adj5 (review* or overview* or synthesis)).tw. (5867)

10 (integrat* adj3 (research or review* or literature)).tw. (8723)

11 (pool* adj2 (analy* or data)).tw. (22510)

12 (handsearch* or (hand adj3 search*)).tw. (7737)

13 (manual* adj3 search*).tw. (4754)

14 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 (374646)

15 ((clinical or care*) adj3 pathway*).ti,ab,kw. (8488)

16 (practice adj3 parameter*).ti,ab,kw. (1186)

17 (guidance* or guideline*).ti. (64813)

18 algorithms/ or clinical protocols/ or critical pathway/ or guidelines as topic/ or practice guidelines as topic/
or Health Planning Guidelines/ or practice guideline/ (417898)

19 Technology Assessment, Biomedical/ (7283)

20 15 or 16 or 17 or 18 or 19 (448362)

21 4 and 14 and 20 (752)

22 limit 21 to english language (719)

Database: MEDLINE in-Process

Results: 63

Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <1946 to August 13, 2019>
Search Strategy:

---

1 (("semi-structured" or semistructured or unstructured or informal or "in-depth" or indepth or "face-to-face" or
structured or guide) adj2 (interview* or discussion* or questionnaire*)).tw,kw. (15764)

2 (focus group* or qualitative or ethnograph* or fieldwork or thematic or phenomenological or "grounded theory" or
"field work" or "key informant").tw,kw. (42243)
3 interviews as topic/ or focus groups/ or narration/ or qualitative research/ (0)
4 1 or 2 or 3 (50295)
5 ("meta-synthesis" or "Meta-synthesis" or "framework synthesis").tw,kw. (204)
6 exp Review Literature as Topic/ (0)
7 (review* or overview*).ti. (77895)
8 (systematic* adj5 (review* or overview*)).tw. (30009)
9 (qualitative* adj5 (review* or overview* or synthesis)).tw. (1472)
10 (integrat* adj3 (research or review* or literature)).tw. (1866)
11 (pool* adj2 (analy* or data)).tw. (3849)
12 (handsearch* or (hand adj3 search*)).tw. (985)
13 (manual* adj3 search*).tw. (826)
14 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 (89196)
15 ((clinical or care*) adj3 pathway*).ti,ab,kw. (1682)
16 (practice adj3 parameter*).ti,ab,kw. (115)
17 (guidance* or guideline*).ti. (10149)
18 algorithms/ or clinical protocols/ or critical pathway/ or guidelines as topic/ or practice guidelines as topic/
or Health Planning Guidelines/ or practice guideline/ (38)
19 Technology Assessment, Biomedical/ (0)
20 15 or 16 or 17 or 18 or 19 (11920)
21 4 and 14 and 20 (64)
22 limit 21 to english language (63)

Database: Embase

Database: Embase <1996 to 2019 Week 32>
Search Strategy:
---

1 (("semi-structured" or semistructured or unstructured or informal or "in-depth" or indepth or "face-to-face" or
structured or guide) adj2 (interview* or discussion* or questionnaire*)).tw,kw. (135921)
2 (focus group* or qualitative or ethnograph* or fieldwork or thematic or phenomenological or "grounded theory" or
"field work" or "key informant").tw,kw. (301375)
3 exp interview/ or narrative/ or qualitative research/ (291630)
4 or/1-3 (531073)
5 (("meta-synthesis" or "Meta-synthesis" or "framework synthesis" or thematic) adj2 analysis).tw,kw. (22818)
6 (review* or overview*).ti. (471369)
7 (systematic* adj5 (review* or overview*)).tw. (204724)
8 (qualitative* adj5 (review* or overview* or synthesis)).tw. (9247)
9 (integrat* adj3 (research or review* or literature)).tw. (12772)
10 (pool* adj2 (analy* or data)).tw. (40771)
11 (handsearch* or (hand adj3 search*)).tw. (10771)
12 (manual* adj3 search*).tw. (7033)
13 or/5-12 (603620)
14 ((clinical or care*) adj3 pathway*).ti,ab,kw. (18202)
15 (practice adj3 parameter*).ti,ab,kw. (1947)
16 (guidance* or guideline*).ti. (102791)
17 algorithm/ or clinical protocols/ or Health care planning/ or practice guideline/ (753032)
18 Biomedical Technology Assessment/ (11507)
19 or/15-18 (796687)
20 4 and 13 and 19 (3821)
21 limit 20 to english language (3727)

Database: CINAHL

Interface - EBSCOhost Research Databases

Search Screen - Advanced Search

Database - CINAHL

#	Query	Results
S4	S1 AND S2 AND S3	884
S3	SYSTEMATIC REVIEW/ OR (review* OR overview*).ti,ab OR (systematic* ADJ5 (review* OR overview*)).ti,ab OR (qualitative* ADJ5 (review* OR overview* OR synthesis)).ti,ab OR (integrat* ADJ3 (research OR review* OR literature)).ti,ab OR (pool* ADJ2 (analy* OR data)).ti,ab OR (handsearch* OR (hand ADJ3 search*)).ti,ab OR (manual* ADJ3 search*).ti,ab	111,849
S2	((("semi-structured" OR semistructured OR unstructured OR informal OR "in-depth" OR indepth OR "face-to-face" OR structured OR guide) ADJ2 (interview* OR discussion* OR questionnaire*)).ti,ab OR (focus group* OR qualitative OR ethnograph* OR fieldwork OR thematic OR phenomenological OR "grounded theory" OR "field work" OR "key informant").ti,ab OR INTERVIEWS/ OR FOCUS GROUPS/ OR NARRATIVES/ OR QUALITATIVE STUDIES/ OR (("meta-synthesis" OR "Meta-synthesis" OR "framework synthesis" OR thematic) ADJ2 analysis).ti,ab)	336,371
S1	(((clinical OR care*) ADJ3 pathway*).ti,ab OR (practice ADJ3 parameter*).ti,ab OR (guidance* OR guideline*).ti,ab OR DECISION SUPPORT TECHNIQUES/ OR PROTOCOLS/ OR PRACTICE GUIDELINES/ OR (Health technology assessment).ti,ab	174,936

Database: PsycINFO

Database: PsycINFO <2002 to August Week 1 2019>

Search Strategy:

---

1 (("semi-structured" or semistructured or unstructured or informal or "in-depth" or indepth or "face-to-face" or
structured or guide) adj2 (interview* or discussion* or questionnaire*)).ti,ab. (76500)

2 (group* or qualitative or ethnograph* or fieldwork or thematic or phenomenological).ti,ab. (672597)

3 "grounded theory".ti,ab. (13384)

4 "field work".ti,ab. (445)

5 "key informant".ti,ab. (1393)

6 Semi-structured interview/ or focus group interview/ or narrative analysis/ or qualitative methods/ (8863)

7 or/1-6 (708637) 8 (("meta-synthesis" or "Meta-synthesis" or "framework synthesis" or thematic) adj2 analysis).ti,ab. (12331) 9 exp Literature Review/ (1076) 10 (review* or overview*).ti. (112644) 11 (systematic* adj5 (review* or overview*)).ti,ab. (29820) 12 (qualitative* adj5 (review* or overview* or synthesis)).ti,ab. (3630) 13 (integrat* adj3 (research or review* or literature)).ti,ab. (8925) 14 (pool* adj2 (analy* or data)).ti,ab. (2906) 15 (handsearch* or (hand adj3 search*)).ti,ab. (1177) 16 (manual* adj3 search*).ti,ab. (905) 17 or/8-16 (144498) 18 ((clinical or care*) adj3 pathway*).ti,ab. (2022) 19 (practice adj3 parameter*).ti,ab. (339) 20 (guidance* or guideline*).ti. (7644) 21 treatment guidelines/ (6190) 22 or/18-21 (13726) 23 7 and 17 and 22 (445) 24 limit 23 to english language (410)

Supplementary Search Techniques:
Google – qualitative synthesis guideline development returned 16 relevant records on the first eight pages. Reference list searching of included papers to be undertaken.

Supplementary file 3: study inclusion checklist

	Author	2000+	QES?	Guidelines?	Methods?	Include?	Refs?

1	Booth 2016	Y	Y	Y	Y	Y	0
2	Campbell 2011	Y	Y	N	Y	N	NA
3	Carroll 2017	Y	Y	Y	Y	Y	0
4	Cowles 2017	Y	N	Y	Y	N	NA
5	DeJean 2016	Y	Y	N	N	N	NA
6	Dixon-Woods 2001	Y	Y	N	Y	N	NA
7	Downe 2019	Y	Y	Y	Y	Y	1
8	Eakin 2003	Y	Y	N	N	N	NA
9	Fadlallah 2019	Y	N	N	Y	N	NA
10	Flemming 2019	Y	Y	Y	Y	Y	1
11	Florez 2018	Y	N	Y	Y	N	NA
12	Gargon 2017	Y	N	Y	N	N	NA
13	Glenton 2016	Y	Y	Y	Y	Y	0
14	Glenton 2019	Y	Y	Y	Y	Y	0
15	Gould 2010	Y	Y	Y	Y	Y	0

16	Grant 2018	Y	N	Y	Y	N	NA
17	Grummer-Strawn 2018	Conference abstract				N	NA
18	Hansen 2011	Y	Y	Y	Y	Y	0
19	Harden 2017	Y	Y	N	Y	N	NA
20	Huls 2018	Conference abstract				N	NA
21	Kelson 2015	Y	N	Y	N	N	NA
22	Knaapen 2015	Y	Y	Y	Y	Y	1
23	Korhonen 2013	Y	Y	N	Y	N	NA
24	Krahn 2008	Y	N	Y	N	N	NA
25	Kristensen 2007	Y	Y	Y	Y	Y	0
26	Langlois 2018	Y	Y	Y	N	N	NA
27	Lewin 2015	Y	Y	Y	N	N	NA
28	Lewin 2018	Y	Y	Y	Y	Y	0
29	Lewin 2019	Y	Y	Y	Y	Y	0
30	Li 2015	Y	N	Y	Y	N	NA

31	Longworth 2011	Y	N	Y	N	N	NA
32	McPherson 2018	Y	N	Y	N	N	NA
33	Murphy 1998	N	N	Y	Y	N	NA
34	NICE Manual 2018	Y	Y	Y	Y	Y	0
35	Noyes 2019	Y	N	Y	Y	N	NA
36	Opiyo 2013	Y	N	Y	Y	N	NA
37	Pope 2002	Y	N	Y	Y	N	NA
38	Ring 2010	Y	Y	Y	Y	Y	1
39	Ring 2011	Y	Y	Y	Y	Y	0
40	Roddis 2018	Y	N	Y	N	N	NA
41	Roddis 2019	Y	Y	Y	N	N	NA
42	Rosedale 2012	Y	N	N	Y	N	NA
43	Saunders 2015	Y	N	N	N	N	NA
44	SBU qual methods Hbk	Y	Y	Y	Y	Y	0
45	Schünemann 2006	Y	N	Y	N	N	NA

46	Staniszewska 2014	Y	N	Y	Y	N	NA
47	Sundberg 2017	Y	N	Y	N	N	NA
48	Tan 2009	Y	Y	Y	Y	Y	0
49	Tong 2014	Y	Y	N	Y	N	NA
50	Van Wesel 2014	Y	Y	N	Y	N	NA
51	Weich 2018	Y	N	Y	N	N	NA
52	Wieringa 2018	Y	N	Y	N	N	NA
53	Zhang 2017	Y	N	Y	N	N	NA

Key *[shaded papers are included in the analysis]*

2000+ - Is the paper published after 2000?

QES – does the paper discuss synthesis or review of qualitative evidence?

Guidelines – does the paper refer specifically to guideline development (or can it be assumed)?

Methods – is there extractable methodological detail?

Refs – number of additional possible includes identified from reference list of paper.

Supplementary file 4: Summary of included papers

Author	Type of paper	Summary
Booth 2016	Guidance	A report funded by the EU as part of a series on evaluating complex interventions ('integrate-HTA'). The guidance document sets out a framework to enable reviewers to choose between different QES

		methods depending on the question they are asking.
Carroll 2017	BMJ Analysis	The analysis focusses on the need for successful guidelines to reflect patient views and argues that qualitative evidence is a key way to do this. The paper is not primarily a detailed methodological paper but contains some extractable methodological detail.
Downe 2019	Research article	The first in a series of three papers that have been written by a group of methodologists working with the WHO on guidelines that integrated QES. The authors examine the use of QES in developing clinical and health systems guidelines.
Flemming 2019	Analysis	This paper presents an overview of the ways QES can be used to address complex interventions.
Glenton 2016	Manual/handbook	Chapter 15 of the WHO handbook for guideline development specifically about using evidence from qualitative research to develop WHO guidelines.
Glenton 2019	Research article	The third in a series of three papers describing the use of QES to inform the development of clinical and health systems guidelines by a team of methodologists who have worked with WHO. The WHO is increasingly using evidence derived from QES to provide information on acceptability and feasibility in its guidelines.
Gould 2010	Methodological report	Gould describes qualitative work done to support the production of two social care guidelines by the UK National Institute for Health and Care Excellence (NICE).

Hansen 2011	Methodological report	This article focuses on qualitative research synthesis in eliciting patients' perspectives as part of the growing drive to include patient views in policy and HTA.
Knaapen 2015	Toolkit (chapter)	Chapter 2 of a GIN toolkit on patient and public involvement in guidelines. It contains practical ideas about how to conduct a qualitative evidence synthesis as part of the guideline development process.
Kristensen 2007	Manual/handbook	The 2007 updated edition of the *Health Technology Assessment Handbook* that was issued by the Danish National Board of Health in 2001 as part of the fulfilment of the National Strategy for HTA. Contains some general detail about QES and also a specific chapter on assessment and syntheses [sic] of qualitative studies (s.4.2).
Lewin 2018	Commentary	Argues that the development of more robust (transparent) methods and tools for QES has widened the opportunities for QES to be used to inform health guidelines (in the context of the WHO).
Lewin 2019	Research article	This is the second in a series of three papers written by methodologists working with the WHO that examines the use of QES in developing clinical and health system guidelines. It specifically discusses using qualitative findings as part of Evidence-to-Decision frameworks.
NICE Manual 2018	Manual	The process manual used by NICE to produce clinical guidelines. The NICE manual includes details

		of synthesis for all the types of evidence it uses, not just qualitative evidence.
Ring 2010	Guidance	Guidance from NHS Quality Improvement Scotland about the various methods of QES that could be used in HTA.
Ring 2011	Research article	The authors conducted a systematic search to identify QES and reflect on the methodological approach used.
Swedish Agency for Health Technology Assessment and Assessment of Social Services (SBU) 2016	Manual	Swedish agency for health technology assessment manual for evaluating and synthesising qualitative material.
Tan 2009	Evidence utilisation report	Describes the use of qualitative research as evidence in a national clinical guideline program (National Institute for Health and Clinical Excellence – NICE, UK).

Supplementary file 5: included papers – full references

Booth, A., Noyes J, Flemming K, Gerhardus, A., Wahlster, P., Van Der Wilt, G.J., Mozygemba, K., Refolo, P., Sacchini, D., Tummers, M., Rehfuess E (2016) Guidance on choosing qualitative evidence synthesis methods for use in health technology assessments of complex interventions.

Carroll, Christopher (2017) Qualitative evidence synthesis to improve implementation of clinical guidelines. BMJ (Clinical research ed.) 356: j80.

Downe, Soo, Finlayson, Kenneth W, Lawrie, Theresa A et al. (2019) Qualitative Evidence Synthesis (QES) for Guidelines: Paper 1 - Using qualitative evidence synthesis to inform guideline scope and develop qualitative findings statements. Health research policy and systems 17(1): 76.

Flemming, Kate, Booth, Andrew, Garside, Ruth et al. (2019) Qualitative evidence synthesis for complex interventions and guideline development: clarification of the purpose, designs and relevant methods. BMJ global health 4(suppl1): e000882.

Glenton C, Lewin S, Lawrie TA et al. (2019) Qualitative Evidence Synthesis (QES) for Guidelines: Paper 3 - Using qualitative evidence syntheses to develop implementation considerations and inform implementation processes. Health research policy and systems 17(1): 74

Glenton, C; Lewin, S; Norris, SL (2016) Using evidence from qualitative research to develop WHO guidelines (Chapter 15). World Health Organization Handbook for Guideline Development. 2nd ed. Geneva: WHO.

Gould, Nick (2010) Integrating qualitative evidence in practice guideline development: Meeting the challenge of evidence-based practice for social work. Qualitative Social Work: Research and Practice 9(1): 93-109.

Hansen H.P.; Draborg E.; Kristensen F.B. (2011) Exploring qualitative research synthesis: The role of patient's perspectives in health policy design and decision making. Patient 4(3): 143-152.

Knaapen, Loes, Colvin, Christopher J, Cowl, Jane et al. How to include qualitative research on patient views in guidelines (2015). GIN Public Toolkit: 28

Kristensen FB & Sigmund H (ed.) (2007) Health Technology Assessment Handbook (https://www.sst.dk/~/media/ECAAC5AA1D6943BEAC96907E03023E22.ashx [accessed 08/09/19]).

Lewin, Simon and Glenton, Claire (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

Lewin, Simon, Glenton, Claire, Lawrie, Theresa A et al. (2019) Qualitative Evidence Synthesis (QES) for Guidelines: Paper 2 - Using qualitative evidence synthesis findings to inform Evidence-to-Decision frameworks and recommendations. Health research policy and systems 17(1): 75

National Institute for Health and Care Excellence (2018) Developing NICE guidelines: the manual.(https://www.nice.org.uk/process/pmg20/chapter/introduction-and-overview [accessed 08/09/19]).

Ring N, Ritchie K, Mandava L JR (2010) A guide to synthesising qualitative research for researchers undertaking health technology assessments and systematic reviews.

Ring, Nicola; Jepson, Ruth; Ritchie, Karen (2011) Methods of synthesizing qualitative research studies for health technology assessment. International journal of technology assessment in health care 27(4): 384-90.

Swedish Agency for Health Technology Assessment and Assessment of Social Services (SBU) (2016) Evaluation and synthesis of studies using qualitative methods of analysis. Stockholm: Swedish Agency for Health Technology Assessment and Assessment of Social Services (SBU); 2016.(https://www.sbu.se/globalassets/ebm/metodbok/sbuhandbook_qualitativemethodsofanalysis.pdf [accessed 08/09/19]).

Tan, Toni P Y; Stokes, Tim; Shaw, Elizabeth J (2009) Use of qualitative research as evidence in the clinical guideline program of the National Institute for Health and Clinical Excellence. International journal of evidence-based healthcare 7(3): 169-72.

Supplementary file 6: papers excluded at full text

Study	Code [Reason]
Campbell R, Pound P, Morgan M et al. (2011) Evaluating meta-ethnography: systematic analysis and synthesis of qualitative research. Health technology assessment (Winchester, England) 15(43): 1-164	- Not about guidelines *[Mentions usefulness to HTA but doesn'tdoesn't actually discuss this.]*
Cowles, Emma, Marsden, Grace, Cole, Amanda et al. (2017) A Review of NICE Methods and Processes Across Health Technology Assessment Programmes: Why the Differences and What is the Impact? Applied health economics and health policy 15(4): 469-477	- Not about QES
DeJean, Deirdre, Giacomini, Mita, Simeonov, Dorina et al. (2016) Finding Qualitative Research Evidence for Health Technology Assessment. Qualitative health research 26(10): 1307-17	- Not about guidelines - No extractable methods
Dixon-Woods M.; Fitzpatrick R.; Roberts K. (2001) Including qualitative research in systematic reviews: Opportunities and problems. Journal of Evaluation in Clinical Practice 7(2): 125-133	- Not about guidelines
Eakin, Joan M and Mykhalovskiy, Eric (2003) Reframing the evaluation of qualitative health research: reflections on a review of appraisal guidelines in the health sciences. Journal of evaluation in clinical practice 9(2): 187-94	- No extractable methods - Not about guidelines
Fadlallah, Racha, El-Jardali, Fadi, Nomier, Mohamed et al. (2019) Using narratives to impact health policy-making: a systematic review. Health Research Policy & Systems 17(1)	- Not about QES - Not about guidelines
Florez, Ivan D, Morgan, Rebecca L, Falavigna, Maicon et al. (2018) Development of rapid guidelines: 2. A qualitative study with WHO guideline developers. Health research policy and systems 16(1): 62	- Not about QES
Gargon, Elizabeth; Williamson, Paula R; Young, Bridget (2017) Improving core outcome set development: qualitative interviews with developers provided pointers to inform guidance. Journal of clinical epidemiology 86: 140-152	- Not about QES - No extractable methods
Grant, Sean, Hazlewood, Glen S, Peay, Holly L et al. (2018) Practical Considerations for Using Online Methods to Engage Patients in Guideline Development. The patient 11(2): 155-166	- Not about QES *[Not about using QES for guideline development.]*

Grummer-Strawn L.M. (2018) Development of evidence-based guidelines at the World Health Organization: A case-study of the ten steps to successful breastfeeding. Breastfeeding Medicine 13(7): a-3	- Conference abstract
Harden, Angela, Thomas, James, Cargo, Margaret et al. (2018) Cochrane Qualitative and Implementation Methods Group guidance series-paper 5: methods for integrating qualitative and implementation evidence within intervention effectiveness reviews. Journal of clinical epidemiology 97: 70-78	- Not about guidelines
Huls S.P., Whichello C., van Exel N.J. et al. (2018) PATIENT PREFERENCES IN HEALTH TECHNOLOGY ASSESSMENT: A SYSTEMATIC REVIEW AND RESEARCH AGENDA. Value in Health 21(supplement3): 325	- Conference abstract
Kelson, Marcia, Akl, Elie A, Bastian, Hilda et al. (2012) Integrating values and consumer involvement in guidelines with the patient at the center: article 8 in Integrating and coordinating efforts in COPD guideline development. An official ATS/ERS workshop report. Proceedings of the American Thoracic Society 9(5): 262-8	- Not about QES - No extractable methods
Korhonen, Anne, Hakulinen-Viitanen, Tuovi, Jylhä, Virpi et al. (2013) Meta-synthesis and evidence-based health care - a method for systematic review. Scandinavian Journal of Caring Sciences 27(4): 1027-1034	- Not about guidelines
Krahn, Murray and Naglie, Gary (2008) The next step in guideline development: incorporating patient preferences. Jama 300(4): 436-438	- No extractable methods - Not about QES
Langlois, Etienne V, Tunçalp, Özge, Norris, Susan L et al. (2018) Qualitative evidence to improve guidelines and health decision making. Bulletin of the World Health Organization 96(2): 79	- No extractable methods
Lewin, Simon, Glenton, Claire, Munthe-Kaas, Heather et al. (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): e1001895	- No extractable methods
Li Y., Yu J., Du L. et al. (2015) Exploration and practice of methods and processes of evidence-based rapid review on peer review of WHO EML application. Journal of Evidence-Based Medicine 8(4): 222-228	- Not about QES
Longworth, Louise, Sculpher, Mark J, Bojke, Laura et al. (2011) Bridging the gap between methods research and the needs of policy makers: a review of the research priorities of the National Institute for Health and Clinical Excellence. International journal of technology assessment in health care 27(2): 180-7	- Not about QES - No extractable methods

McPherson S., Rost F., Town J. et al. (2018) Epistemological flaws in NICE review methodology and its impact on recommendations for psychodynamic psychotherapies for complex and persistent depression. Psychoanalytic Psychotherapy 32(2): 102-121	- Not about QES - No extractable methods
Murphy, E, Dingwall, R, Greatbatch, D et al. (1998) Qualitative research methods in health technology assessment: a review of the literature. Health technology assessment (Winchester, England) 2(16): iii-274	- Pre-2000 - Not about QES
Noyes, Jane, Booth, Andrew, Moore, Graham et al. (2019) Synthesising quantitative and qualitative evidence to inform guidelines on complex interventions: clarifying the purposes, designs and outlining some methods. BMJ global health 4(suppl1): e000893	- Not about QES
Opiyo, Newton, Shepperd, Sasha, Musila, Nyokabi et al. (2013) Comparison of alternative evidence summary and presentation formats in clinical guideline development: a mixed method study. PloS one 8(1): e55067	- Not about QES
Pope, Catherine; Van Royen, Paul; Baker, Richard (2002) Qualitative methods in research on healthcare quality. BMJ Quality & Safety 11(2): 148-152	- Not about QES
Richter Sundberg, Linda; Garvare, Rickard; Nystrom, Monica Elisabeth (2017) Reaching beyond the review of research evidence: a qualitative study of decision making during the development of clinical practice guidelines for disease prevention in healthcare. BMC health services research 17(1): 344	- Not about QES - No extractable methods
Ring N, Ritchie K, Mandava L JR (2010) A guide to synthesising qualitative research for researchers undertaking health technology assessments and systematic reviews	- Duplicate reference
Roddis, Jennifer Karen, Liversedge, Hannah L, Ryder, Isobel et al. (2018) Incorporating the patient experience into clinical guidelines: recommendations for researchers and guideline developers. BMJ evidence-based medicine: bmjebm-2018	- No extractable methods
Rosedale, Mary, Malaspina, Dolores, Malamud, Daniel et al. (2012) Developing patient-centered treatment protocols in brain stimulation: a rationale for combining quantitative and qualitative approaches in persons with HIV. Journal of the American Psychiatric Nurses Association 18(3): 166-74	- Not about QES - Not about guidelines
Saunders, Hannele (2015) Translating knowledge into best practice care bundles: a pragmatic strategy for EBP implementation via moving postprocedural pain management nursing guidelines into clinical practice. Journal of clinical nursing 24(1314): 2035-51	- Not about QES - Not about guidelines - No extractable methods

Schünemann, Holger J; Fretheim, Atle; Oxman, Andrew D (2006) Improving the use of research evidence in guideline development: 10. Integrating values and consumer involvement. Health research policy and systems 4(1): 22	- Not about QES
Staniszewska, Sophie, Boardman, Felicity, Gunn, Lee et al. (2014) The Warwick Patient Experiences Framework: patient-based evidence in clinical guidelines. International Journal for Quality in Health Care: journal of the International Society for Quality in Health Care 26(2): 151-7	- Not about QES
Tong A., Palmer S., Craig J.C. et al. (2016) A guide to reading and using systematic reviews of qualitative research. Nephrology Dialysis Transplantation 31(6): 897-903	- Not about guidelines
van Wesel, Floryt, Alisic, Eva, Boeije, Hennie et al. (2014) Using qualitative evidence to optimize child PTSD treatment guidelines. Psychological Trauma: Theory, Research, Practice, and Policy 6(5): 546-554	- Not about guidelines
Weich S, Fenton SH, Bhui K et al. Realist Evaluation of the Use of Patient Experience Data to Improve the Quality of Inpatient Mental Health Care (EURIPIDES) in England: study protocol. BMJ open 8(6): e021013	- No extractable methods - Not about QES
Wieringa, Sietse, Dreesens, Dunja, Forland, Frode et al. (2018) Different knowledge, different styles of reasoning: a challenge for guideline development. BMJ evidence-based medicine 23(3): 87-91	- Not about QES - No extractable methods
Zhang, Yuan, Coello, Pablo Alonso, Brozek, Jan et al. (2017) Using patient values and preferences to inform the importance of health outcomes in practice guideline development following the GRADE approach. Health and quality of life outcomes 15(1): 52	- Not about QES - No extractable methods

APPENDIX F: ENTREQ CRITERIA, DESCRIPTIONS AND CODING RULES

Table 17: Coding rules for specific ENTREQ criteria

Item	Guide and description	Specific rules for this analysis. Criteria were marked as met if the QES did the following
Aim	State the research question the synthesis addresses.	Stated the review question to be addressed.
Title & Synthesis methodology	Identify the synthesis methodology or theoretical framework which underpins the synthesis and describe the rationale for choice of methodology (e.g. meta-ethnography, thematic synthesis, critical interpretive synthesis, grounded theory synthesis, realist synthesis, meta-aggregation, meta-study, framework synthesis).	Specified a methodology beyond noting that the methods set out in the NICE guideline manual had been followed or that 'thematic synthesis was undertaken'. Detailed exposition of methods was not required.
Approach to searching	Indicate whether the search was pre-planned (comprehensive search strategies to seek all available studies) or iterative (to seek all available concepts until theoretical saturation is achieved).	No additional interpretation.
Inclusion criteria	Specify the inclusion/exclusion criteria (e.g. in terms of population, language, year limits, type of publication, study type).	No additional interpretation.
Data sources	Describe the information sources used (e.g. electronic databases (MEDLINE, EMBASE, CINAHL,	No additional interpretation.

	PsychINFO, Econlit), grey literature databases (digital thesis, policy reports), relevant organisational websites, experts, information specialists, generic web searches (Google Scholar), hand-searching, reference lists) and when the searches were conducted; provide the rationale for using the data sources.	
Electronic Search strategy	Describe the literature search (e.g. provide electronic search strategies with population terms, clinical or health topic terms, experiential or social phenomena related terms, filters for qualitative research and search limits).	No additional interpretation.
Study screening methods	Describe the process of study screening and sifting (e.g. title, abstract and full text review, number of independent reviewers who screened studies).	Description of process of sifting titles and abstracts then full text review sufficient for mark.
Study characteristics	Present the characteristics of the included studies (e.g. year of publication, country, population, number of participants, data collection, methodology, analysis, research questions).	If evidence tables were present in the review, then this was considered to meet this criterion.
Study selection results	Identify the number of studies screened and provide reasons for study exclusion (e.g. for comprehensive searching, provide numbers of studies screened and reasons for exclusion indicated in a	Presence of a PRISMA style flowchart for study selection was considered sufficient for this criterion.

	figure/flowchart; for iterative searching describe reasons for study exclusion and inclusion based on modifications to the research question and/or contribution to theory development).	
Rationale for appraisal	Describe the rationale and approach used to appraise the included studies or selected findings (e.g. assessment of conduct (validity and robustness), assessment of reporting (transparency), assessment of content and utility of the findings).	No additional interpretation.
Appraisal items	State the tools, frameworks and criteria used to appraise the studies or selected findings (e.g. Existing tools: CASP, QARI, COREQ; reviewer developed tools; describe the domains assessed: research team, study design, data analysis and interpretations, reporting).	Statement of critical appraisal tool used, or description of alternative critical appraisal process considered sufficient.
Appraisal process	Indicate whether the appraisal was conducted independently by more than one reviewer and if consensus was required.	No additional interpretation.
Appraisal results	Present results of the quality assessment and indicate which articles, if any, were weighted/excluded based on the assessment and give the rationale.	Presence of critical appraisal assessment in evidence table or summary of studies table considered sufficient.
Data extraction	Indicate which sections of the primary studies were analysed and how were the data extracted from the primary studies? (e.g. all text	No additional interpretation.

	under the headings “results /conclusions” were extracted electronically and entered into a computer software).	
Software	State the computer software used, if any.	No additional interpretation.
Number of reviewers	Identify who was involved in coding and analysis.	No additional interpretation.
Coding	Describe the process for coding of data (e.g. line by line coding to search for concepts).	No additional interpretation.
Study comparison	Describe how were comparisons made within and across studies (e.g. subsequent studies were coded into pre-existing concepts, and new concepts were created when deemed necessary).	Presence of a logic model, theme map or similar was considered sufficient.
Derivation of themes	Explain whether the process of deriving the themes or constructs was inductive or deductive.	No additional interpretation.
Quotations	Provide quotations from the primary studies to illustrate themes/constructs and identify whether the quotations were participant quotations or the author’s interpretation.	No additional interpretation.
Synthesis output	Present rich, compelling and useful results that go beyond a summary of the primary studies (e.g. new interpretation, models of evidence, conceptual models, analytical framework, development of a new theory or construct).	No additional interpretation.

APPENDIX G: R CODE USED TO GENERATE FIGURES

R code (RStudio version 1.3.1056) used to generate figures 6, 8 and 9. Other figures were generated in Microsoft Excel (Microsoft Office Professional Plus 2019).

Figure 3, **Figure 4** and **Figure 5** were produced in excel from the data spreadsheet using inbuilt excel functions.

For Figure 6 the following code was used on the 'mean ENTREQ by centre/year' supplementary spreadsheet

```
library(ggplot2)

library(readxl)

#mean ENTREQ score by year across different centres

mean_entreq_by_year <- read_excel("mean entreq by year.xlsx",

+ col_types = c("character", "numeric", "numeric"))

View(mean_entreq_by_year)

df<- mean_entreq_by_year

ggplot( df, aes( x = Year, y = Mean ) ) +

geom_bar( stat = "identity" ) +

facet_wrap( ~ Centre ) +

ylab("mean ENTREQ score out of 21")
```

For **Figure 7** the following code was used on the 'clean data' supplementary spreadsheet available at https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-022-01743-1#data-availability

```
library(tidyverse)
```

```
df<-Clean_data_for_R_v_1_0

 # plot of median ENTREQ score by authoring centre

At<- ggplot(df, aes(x=Author, y=Total))+

 geom_boxplot()

At+

 geom_jitter(shape=16, position=position_jitter(0.2))+

 labs(x="Author of QES", y="Total ENTREQ score", title= "Median ENTREQ score by author of QES", subtitle= "Dots represent individual scores")

#End
```

APPENDIX H: CONSOLIDATED CRITERIA FOR REPORTING QUALITATIVE STUDIES (COREQ): 32-ITEM CHECKLIST

This checklist pertains to the primary qualitative study reported in Chapters 4 and 5.

Developed from:

Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. International Journal for Quality in Health Care. 2007. Volume 19, Number 6: pp. 349 – 357

Table 18: COREQ checklist

No. Item	Guide questions/description	Reported in (Chapter and section)
Domain 1: Research team and reflexivity		
Personal characteristics		
1. Interviewer/facilitator	Which author/s conducted the interview or focus group?	Implicit – PhD thesis
2. Credentials	What were the researcher's credentials? E.g. PhD, MD	Implicit – PhD thesis
3. Occupation	What was their occupation at the time of the study?	Ch 4, reflexivity statement, Personal background and positioning
4. Gender	Was the researcher male or female?	Ch 4, reflexivity statement, Personal

		background and positioning
5. Experience and training	What experience or training did the researcher have?	Ch 1 Introduction Ch 4, reflexivity statement, Personal background and positioning.
Relationship with participants		
6. Relationship established	Was a relationship established prior to study commencement?	Ch 4- Interactions with participants
7. Participant knowledge of the interviewer	What did the participants know about the researcher? e.g. personal goals, reasons for doing the research	Ch 4- Interactions with participants. Participant in formation sheet (appendices)
8. Interviewer characteristics	What characteristics were reported about the interviewer/facilitator? e.g. Bias, assumptions, reasons and interests in the research topic	Ch 4 – reflexivity section

Domain 2: study design		
Theoretical framework		
9. Methodological orientation and Theory	What methodological orientation was stated to underpin the study? e.g. grounded theory, discourse analysis, ethnography, phenomenology, content analysis	Ch 4 -Background to the qualitative study Ch 4 – research design
Participant selection		
10. Sampling	How were participants selected? e.g. purposive, convenience, consecutive, snowball	Ch 4 - participants
11. Method of approach	How were participants approached? e.g. face-to-face, telephone, mail, email	Ch 4 – recruitment process
12. Sample size	How many participants were in the study?	Ch 5 - participants
13. Non-participation	How many people refused to participate or dropped out? Reasons?	Ch 5 - participants
Setting		
14. Setting of data collection	Where was the data collected? e.g. home, clinic, workplace	Ch 4 – data collection methods
15. Presence of non-participants	Was anyone else present besides the participants and researchers?	Ch 4 – data collection methods
16. Description of sample	What are the important characteristics of the sample? e.g. demographic data, date	Ch 4 – participants, Ch 5 - participants
Data collection		

17. Interview guide	Were questions, prompts, guides provided by the authors? Was it pilot tested?	Ch 4 – Topic guide
18. Repeat interviews	Were repeat interviews carried out? If yes, how many?	Ch 5 - Participants
19. Audio/visual recording	Did the research use audio or visual recording to collect the data?	Ch 4 – data recording and transcription
20. Field notes	Were field notes made during and/or after the interview or focus group?	Ch 4 – data recording and transcription
21. Duration	What was the duration of the interviews or focus group?	Ch 5 - Participants
22. Data saturation	Was data saturation discussed?	Ch 4 – research design
23. Transcripts returned	Were transcripts returned to participants for comment and/or correction?	Ch 5 - Participants
Domain 3: analysis and findings		
Data analysis		
24. Number of data coders	How many data coders coded the data?	Implicit – PhD thesis
25. Description of the coding tree	Did authors provide a description of the coding tree?	Ch 4 – data analysis
26. Derivation of themes	Were themes identified in advance or derived from the data?	Ch 4 – data analysis
27. Software	What software, if applicable, was used to manage the data?	Ch 4 – data analysis

28. Participant checking	Did participants provide feedback on the findings?	Ch 5 - Participants
Reporting		
29. Quotations presented	Were participant quotations presented to illustrate the themes/findings? Was each quotation identified? e.g. participant number	Ch 5 throughout
30. Data and findings consistent	Was there consistency between the data presented and the findings?	Ch 5 throughout
31. Clarity of major themes	Were major themes clearly presented in the findings?	Ch 5 throughout
32. Clarity of minor themes	Is there a description of diverse cases or discussion of minor themes?	Ch 5 throughout

APPENDIX I: PARTICIPANT INFORMATION SHEET

Participant Information Sheet [24/02/2021]

Research project: Using qualitative research to make guideline recommendations

You are being invited to take part in a research project. Before you decide whether or not to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read this information sheet carefully and discuss it with others if you want. Ask me if you have any questions.

What is the research about?

NICE guideline committees use different kinds of evidence to make decisions about what they want to recommend. One of the kinds of evidence is a qualitative evidence synthesis (that draws together and summarises evidence from qualitative research studies). This research is interested in what part (if any) qualitative evidence synthesis plays in informing decisions about recommendations.

1. Why are you asking me?

- I am asking people on committees that are looking at research evidence which includes qualitative evidence syntheses whether I can observe a meeting where you are in attendance.
- In addition, I am asking a small sample of people from each of the committees to take part in an online interview to talk about their experience of using qualitative

evidence on the committee, whether it was useful or not, and if and how it influenced the recommendations made. I will be interviewing 12 people in total.

- I am seeking to interview clinicians, lay-people and other professionals. People will be chosen by role to maximise the diversity of viewpoints in the sample.

2. Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep (and be asked to sign a consent form).

It will not be possible to withdraw from the observation once it has taken place. This is because the notes I take will not identify you or your specific contribution.

If you have been interviewed and later wish to withdraw from the research, please contact me (details below).

If you wish to withdraw, please let me know as soon as possible (within one month) because if I have already analysed the data, I can't take out anything you have told me.

3. What do I have to do?

If I am observing a committee meeting that you are attending then you do not have to do anything more than you would normally do at a committee meeting. I will have my camera and microphone switched off and will not take part in the meeting at all. I will be making notes about the way the committee discuss the evidence and use it in recommendations. You will be asked to formally consent to this prior to the committee meeting.

I am also trying to find a few people (4 people from each committee) to be interviewed. If you tick the box on your consent form that you would be willing to be interviewed then I may ask you for an interview. Most committee members will not be asked for an interview as I only need a small number of people. If I do take you up on your offer to be interviewed then I will email you so that we can arrange a mutually convenient time for us to conduct the interview. The interview will be done online via Google Meet. I can help you to use it if you have not done so before. It does not require you to download or install any software onto your computer or smartphone.

During the interview we will have a discussion about your experience of using qualitative research in committee meetings. I will send you a list of the topics that I would like us to discuss in advance of the interview so you have plenty of time to think about it, but you will also have plenty of opportunity to tell me about anything else that you think is important.

- The interview will last around 45 minutes to 1 hour.

I am only interested in your experience of using qualitative evidence in the committee meeting. I won't ask you any questions about the committee and your views of how it is run, or about any other members.

4. Is there any risk from taking part?

I do not expect there to be any risks from taking part. If you are a lay-member, and take part in an interview, it is possible that talking about committee decisions that relate to a condition that you have experienced or are living with might upset you. If you become upset I will stop and ask if you would like a break, or to stop the interview.

5. Are there benefits of taking part?

I hope that this research will help to understand how findings from qualitative research can be best taken into account in committees that make recommendations about health and social care.

8. Will anyone know I took part?

All the information that I collect about you during the course of the research will be kept strictly confidential and will only be accessible to myself and in anonymised form to my two supervisors. You will not be able to be identified in any reports or publications.

9. What is the legal basis for processing my personal data?

According to data protection legislation, I am required to inform you that the legal basis I am applying in order to process your personal data is that "processing is necessary for the performance of a task carried out in the public interest" (Article 6(1)(e)). Further

information can be found in the University's Privacy Notice https://www.sheffield.ac.uk/govern/data-protection/privacy/general.

10. What will happen to the data collected? What will happen to the results of the project?

The notes that I take during the committee meetings will be scanned as.pdf documents and uploaded to a secure University storage folder which is password protected and only accessible to myself and my supervisors. The paper version will be deleted immediately following transfer.

The recordings of the interviews will be uploaded to a secure University storage folder which is password protected and only accessible to myself and my supervisors. The version stored in Google Meet will be deleted immediately following transfer. The typed-up transcripts of the recordings will be made anonymous by taking out any references to your name or any place or condition that might make you identifiable. These transcripts will be used to provide short quotes in my thesis, future research papers, and for illustration in conference presentations and lectures. No other use will be made of them without your written permission.

All the data will be destroyed three years after collection, when I finish my PhD, whichever is soonest.

My findings will be written into a PhD thesis at the University of Sheffield. This will be freely available for people to read.

11. Who is organising and funding the research?

I am funding my own PhD. No funding is coming from any other source.

12. Who is the Data Controller?

The University of Sheffield will act as the Data Controller for this study. This means that the University is responsible for looking after your information and using it properly.

13. Who has ethically reviewed the project?

This project has been ethically reviewed and approved by the School of Health and Related Research Ethics Committee at The University of Sheffield.

14. What if something goes wrong and I wish to complain about the research?

If you want to complain about the way the study has been carried out, please contact my primary supervisor Dr Susan Baxter s.k.baxter@sheffield.ac.uk 0114 2222436 in the first instance.

If you feel your complaint has not been handled properly, you can contact the Head of Department in the School for Health and Related Research.

Professor John Brazier j.e.brazier@sheffield.ac.uk

If your complaint relates to how your personal data has been handled, information about how to raise a complaint can be found in the University's Privacy Notice: https://www.sheffield.ac.uk/govern/data-protection/privacy/general.

If you wish to contact the Data Protection Officer at the University please write to: Luke Thompson, Head of Data Protection and Legal Services, University Secretary's Office, University of Sheffield, Western Bank, Sheffield, S10 2TN or email dataprotection@sheffield.ac.uk

15. Contact for further information

If you need further information about the study, please contact

Chris Carmona ccarmona1@sheffield.ac.uk 0207 045 2155

APPENDIX J: CONSENT FORM

Qualitative evidence synthesis in guideline development: Consent form

Please tick the appropriate boxes	**Yes**	**No**
Taking part in the project		
I have read and understood the project information sheet dated DD/MM/YYYY or the project has been fully explained to me. (If you will answer No to this question please do not proceed with this consent form until you are fully aware of what your participation in the project will mean.)	☐	☐
I have been given the opportunity to ask questions about the project.	☐	☐
I agree to take part in the project. I understand that taking part in the project will participating in a semi-structured interview about my experiences of using qualitative evidence synthesis to make recommendations on a NICE guideline committee. The interview will last from 45 mins to 1 hour.	☐	☐
I understand that my taking part is voluntary and that I can withdraw from the study at any time until a month after my interview; I do not have to give any reasons for why I no longer want to take part and there will be no adverse consequences if I choose to withdraw.	☐	☐
How my information will be used during and after the project		
I understand my personal details such as name, phone number, address and email address etc. will not be revealed to people outside the project.	☐	☐
I understand and agree that my words may be quoted in publications, reports, web pages, and other research outputs. I understand that I will not be named in these outputs.	☐	☐
I understand that the security of my telephone or computer device that I use for my interview is my responsibility and that if I am concerned about security then I should not say anything that I would not otherwise say.	☐	☐
So that the information you provide can be used legally by the researchers		
I agree to assign the copyright I hold in any materials generated as part of this project to The University of Sheffield.	☐	☐

Name of participant [printed] Signature Date

Name of Researcher [printed] Signature Date

Chris Carmona

Project contact details for further information:

Researcher: Chris Carmona ccarmona1@sheffield.ac.uk 0207 045 2155

APPENDIX K: LETTER OF ETHICAL APPROVAL FOR QUALITATIVE STUDY

The University Of Sheffield.

Downloaded: 31/07/2020
Approved: 31/07/2020

Christopher Carmona
Registration number: 180272761
School of Health and Related Research
Programme: PhD Health and Related Research

Dear Christopher

PROJECT TITLE: Qualitative Evidence Synthesis in guideline development
APPLICATION: Reference Number 035876

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 31/07/2020 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 035876 (form submission date: 30/07/2020); (expected project end date: 31/08/2021).
- Participant information sheet 1081335 version 3 (29/07/2020).
- Participant information sheet 1081336 version 1 (02/07/2020).
- Participant consent form 1081337 version 2 (29/07/2020).
- Participant consent form 1081338 version 2 (29/07/2020).

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Your responsibilities in delivering this research project are set out at the end of this letter.

Yours sincerely

Jennifer Burr
Ethics Administrator
School of Health and Related Research

Please note the following responsibilities of the researcher in delivering the research project:

- The project must abide by the University's Research Ethics Policy: https://www.sheffield.ac.uk/rs/ethicsandintegrity/ethicspolicy/approval-procedure
- The project must abide by the University's Good Research & Innovation Practices Policy: https://www.sheffield.ac.uk/polopoly_fs/1.671066!/file/GRIPPolicy.pdf
- The researcher must inform their supervisor (in the case of a student) or Ethics Administrator (in the case of a member of staff) of any significant changes to the project or the approved documentation.
- The researcher must comply with the requirements of the law and relevant guidelines relating to security and confidentiality of personal data.
- The researcher is responsible for effectively managing the data collected both during and after the end of the project in line with best practice, and any relevant legislative, regulatory or contractual requirements.

APPENDIX L: THE MOVE TOWARDS LIVING SYSTEMATIC REVIEWS AND LIVING GUIDELINES IN HEALTHCARE: CONSIDERATION OF THE POSSIBILITIES AND CHALLENGES FOR LIVING QUALITATIVE EVIDENCE SYNTHESES

Carmona, C., Carroll, C. & Baxter, S. The move towards living systematic reviews and living guidelines in healthcare: consideration of the possibilities and challenges for living qualitative evidence syntheses. *Syst Rev* **12**, 47 (2023). https://doi.org/10.1186/s13643-023-02218-0

Chris Carmona MSc, School of Health and Related Research (ScHARR), University of Sheffield, UK

Christopher Carroll PhD, School of Health and Related Research (ScHARR), University of Sheffield, UK

Susan Baxter PhD, School of Health and Related Research (ScHARR), University of Sheffield, UK

Abstract

Over the past decade QES, a range of methods for synthesising qualitative research evidence, has become a valued form of evidence for guideline producers who wish to understand more about patient preference and acceptability of treatments. The surge in interest in living systematic reviews and the appearance of living guidelines as a response to the COVID-19 pandemic potentially weakens the value and usability of QES.

There are currently no published methods for producing living QES, and If QES are to remain of worth to guideline producers then methods for the rapid, frequent updating of them will need to be developed. We discuss some of the similarities and differences between qualitative and quantitative evidence syntheses and highlight areas where development is needed if reviewers are to progress living approaches to QES.

Keywords: Qualitative Evidence Synthesis, health guidelines, living guidelines, living systematic review.

Background

QES refers to a range of methods for synthesising qualitative research studies, and has been in use since the late 1980s(1). Since 2004, there has been a Cochrane Methods Group tasked with advising the Cochrane Collaboration on policy related to the synthesis of qualitative evidence and the integration of qualitative evidence with Cochrane effectiveness reviews. More recently, QES have become a part of the process of developing evidence-based health guidelines by organisations such as the UK National Institute for Health and Care Excellence (NICE) and by the WHO where they have been used by guideline panels to support their

decision making.(2-4) This incorporation of QES into health guidelines has been made easier both by methodological developments in the ways that QES are undertaken (for example the introduction of GRADE-CERQual(5)), and a drive by guideline-producing organisations to consider the effects of patient preference, feasibility, and acceptability on the broader effectiveness of a treatment or intervention when making guideline recommendations.(3)

The concept of a living systematic review (LSR) has been in evidence since 2014.(6) Cochrane define a LSR as a "*systematic review which is continually updated, incorporating relevant new evidence as it becomes available."*(7) Prior to the COVID-19 pandemic LSRs were largely theoretical entities, although Cochrane released their first 'Guidance for the production and publication of Cochrane living systematic reviews' in 2017 (and updated it in 2019).(8) During the COVID pandemic, the need to respond quickly to rapidly changing evidence and practice led to the use of LSRs to inform living guidelines (LG), that could be updated each time the evidence changed in a meaningful way. A living guideline is "an optimisation of the guideline development process to allow updating of individual recommendations as soon as new relevant evidence becomes available."(9)

This surge in interest in LGs has been facilitated by developments in the technologies used to search for and screen evidence using machine learning, and also by developments of user-friendly updatable content management systems such as MAGICapp.(10) Living guidelines are also a response to the need for clinicians and healthcare professionals to have access to guidelines based on the best currently available evidence.

The value of producing LGs has been recognised by the Australian National Clinical Evidence Taskforce (NCET), NICE in the UK and the WHO. The WHO has a living guideline for the clinical management of COVID-19(11), NCET has living guidelines on both COVID-19(12) and mpox(13), and pillar 2 of the NICE strategy 2021- 2026 promises "Dynamic, living guideline recommendations" over the next 3 years.(14)

What are the implications of LGs for the future use of QES in health guideline development?

The increase in focus on LGs based on living systematic reviews poses a challenge to the 'new era for qualitative research' described by Lewis and Glenton.(15) Currently there are no published methods for constant updating of QES, or for the development of 'living QES.' If qualitative methodologists cannot respond to the challenge of developing methods for constantly updating QES then they run the risk of being side-lined by a focus on quantitative evidence and syntheses for LGs. Qualitative methodologists urgently need to develop methods for making QES 'living' in a way that will allow them to be updated alongside LSR.

Characteristics of living reviews and guidelines

What makes a systematic review 'living'?

Cochrane (7) define a LSR as a systematic review which is continually updated, incorporating relevant new evidence as it becomes available.

They define this in practical terms as LSRs:

being underpinned by continual, active monitoring of the evidence (i.e., monthly searches)

immediately including any new important evidence (meaning data, studies, or information) that is identified

being supported by up-to-date communication about the status of the review, and any new evidence being incorporated

including pre-defined decisions about how often new evidence is sought and screened and when and how new evidence is incorporated into the review

If the review is set up in the right way from its inception then this can be fairly straightforward. For a review that is authored within a program or application that automatically undertakes analysis, for example Cochrane's RevMan(16) or MAGICapp(10), then new data can be added to an existing meta-analysis to generate a new pooled effect estimate, and GRADE (if it is being used) domains for that outcome can be edited if necessary. Lists of included and excluded studies can be updated and new evidence tables inserted.

What makes a health guideline 'living'?

It currently seems less clear what the criteria are for a LG. Akls definition (above)(9) provides a good starting point, however the processes that need to underpin the guidelines production and maintenance remain exploratory. It is broadly agreed that a process needs to be established whereby new evidence that updates a LSR is assessed in relation to existing guideline recommendations, and then a judgment is made about whether the new evidence is likely to affect the existing recommendation. If it is, then a guideline panel will meet to discuss the evidence and update the recommendation, if not the LSR will be updated but the guideline will stand.

How can we apply these characteristics to QES?

Towards a living QES

If we apply the list of criteria from LSRs to QES, one can imagine a scenario where searches for qualitative publications are repeated regularly to identify new studies for inclusion into a QES. Searching for qualitative studies however is typically perceived to be more complex than searching for randomised controlled studies.(17) Numbers of records retrieved can be higher, which in turn implies more work for researchers tasked with frequent sifting of this data. It might be that monthly updates are considered too frequent for most areas of health reviews given that fewer qualitative studies are published.(18) The main challenge for a

living QES lies in the ability of the QES to incorporate new evidence quickly and meaningfully in a way that allows decisions to be made by guideline producers about whether a guideline needs updating.

We might suppose that different methods of qualitative evidence synthesis will encounter distinct levels of challenge in attempting to integrate new data into existing themes. It seems likely that the more interpretive approaches to QES, for example meta-ethnography(1), might require substantial work to incorporate even modest amounts of new data because large parts of the analysis might need to be re-done to take account of the new data. Conversely, one might imagine that for more aggregative approaches to QES it will be more straightforward to determine the potential impact of new data on existing themes. This might be the case, for example, for aggregative synthesis(19) or approaches that use pre-specified frameworks for synthesising data, for example best-fit framework synthesis.(20) If the data match closely to what is already contained in the theme, then the theme may not need updating, other than to add the study details to the review and consider the effects of the study on CERQual decisions (if used) for the theme. Changes to themes might be easily integrated if they require refinement of individual themes or aggregates of codes rather than a wholescale reinterpretation of the data.

There are further considerations that may be unique to QES that need discussion and development, stemming from the very different nature of the evidence used to develop QES. A good example of this is the currency of qualitative data (its up-to-date-ness). While we can probably assume that in the context of an LSR, evidence about the efficacy of a drug is likely to be constant over time, we might not be able to say the same of qualitative data. Prevalent social and individual views and experiences change over time as society, health care, and health expectations change over time. There might be a requirement for evidence to be removed from a QES as it becomes dated. This raises questions regarding at what point should evidence be ‘retired’ from a living QES? Is there a lifespan for a QES before it becomes incoherent and needs to be completely revised?

How will such a living QES inform a living health guideline?

In the same way as for a LSR, a guideline producer using a living QES to inform part of a health guideline would need explicit criteria to invest the time and resource necessary to convene a guideline panel to re-examine existing recommendations on the basis of integrating new evidence. There would need to be some belief or expectation that the new evidence would change recommendations. While for quantitative reviews decision making may be based around whether new evidence changes the effect size (or direction of effect) of an intervention, for qualitative data the implications of adding new studies may be less clear. Qualitative researchers may not be comfortable with the idea of ‘hard’ rules about updating guidelines. Perhaps a meaningful alternative would be to consult a small panel of experts and lay-people each time a QES is updated to seek guidance on whether new evidence has potential to alter existing recommendations.

Conclusion

The past 10 years have been a period of growth in methods for QES with their value increasingly recognised by organisations that specialise in evidence-based medicine. The emergence of LSRs and LGs requires an urgent response from QES methodologists to develop efficient and effective processes for updating QES quickly and frequently, if the synthesis of evidence from qualitative studies is to meet the needs of health guideline producers.

List of abbreviations

COVID-19	Corona Virus Disease 2019
GRADE	Grading of Recommendations Assessment, Development and Evaluation
CERQual	Confidence in the Evidence from Reviews of Qualitative Research
LG	Living Guideline
LSR	Living Systematic Review
NCET	Australian National Clinical Evidence Taskforce
NICE	National Institute for Health and Care Excellence (UK)
QES	Qualitative Evidence Synthesis
WHO	World Health Organization

Declarations

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable

Availability of data and materials: Not applicable

Competing interests: Chris Carmona is an employee of NICE. Susan Baxter and Christopher Carroll have no competing interests.

Funding: No funding was received to write this commentary. Open access publication of this paper was funded by the University of Sheffield Institutional Open Access Fund.

Authors' contributions: Chris Carmona drafted the commentary, Susan Baxter and Chris Carroll commented on the manuscript and helped to develop the arguments. All authors read and approved the final manuscript.

Acknowledgements: None

References

1. Noblit GW, Hare RD. *Meta-ethnography: Synthesizing qualitative studies*. Newbury Park, Calif: Sage Publications.; 1988.

2. Flemming K, Booth A, Garside R, Tuncalp O, Noyes J. Qualitative evidence synthesis for complex interventions and guideline development: clarification of the purpose, designs and relevant methods. BMJ Global Health. 2019;4.

3. Carmona C, Baxter S, Carroll C. Systematic review of the methodological literature for integrating qualitative evidence syntheses into health guideline development. Research Synthesis Methods. 2021.

4. Carmona C, Baxter S, Carroll C. The conduct and reporting of qualitative evidence syntheses in health and social care guidelines: a content analysis. BMC Med Res Methodol. 2022;22(1):267.

5. GRADE-CERQual: Confidence in the Evidence from Reviews of Qualitative Research [Available from: https://www.cerqual.org/].

6. Elliott JH, Turner T, Clavisi O, Thomas J, Higgins JP, Mavergames C, et al. Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. PLoS Med. 2014;11(2):e1001603.

7. Cochrane Community. Living Systematic Reviews [Available from: https://community.cochrane.org/review-production/production-resources/living-systematic-reviews].

8. Brooker J, Synnot A, McDonald S, Elliott J, Turner T. Cochrane living systematic reviews: Interim guidance for pilots (Version 1.0). 2017. Contract No.: 29/01/2023.

9. Akl EA, Meerpohl JJ, Elliott J, Kahale LA, Schünemann HJ, Agoritsas T, et al. Living systematic reviews: 4. Living guideline recommendations. Journal of Clinical Epidemiology. 2017;91:47-53.

10. MAGIC. MAGICapp.

11. World Health Organization. Living guidance for clinical management of COVID-19. 2021. Report No.: WHO/2019-nCoV/clinical/2021.2.

12. National Clinical Evidence Taskforce. Australian guidelines for the clinical care of people with COVID-19 (Version 70.1) 2022.

13. National Clinical Evidence Taskforce. Australian guidelines for the clinical care of people with MPX (Version 4.1). 2022.

14. National Institute for Health and Care Excellence (NICE). NICE strategy 2021 to 2026: Dynamic, Collaborative, Excellent. Manchester, UK: National Institute for Health and Care Excellence; 2021.

15. Lewin S, Glenton C. 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. 2018;17.

16. The Cochrane Collaboration. Review Manager (RevMan). 5.4 ed2020.

17. Booth A. Searching for qualitative research for inclusion in systematic reviews: a structured methodological review. Syst Rev. 2016;5:74.

18. DeJean D, Giacomini M, Simeonov D, Smith A. Finding Qualitative Research Evidence for Health Technology Assessment. Qualitative Health Research. 2016;26(10):1307-17.

19. Lockwood C, Porritt K, Munn Z, Rittenmeyer L, Salmond S, Bjerrum M, et al. Systematic reviews of qualitative evidence. In: E A, Z M, editors. JBI Manual for Evidence Synthesis 2020.

20. Carroll C, Booth A, Cooper K. A worked example of "best-fit" framework synthesis: A systematic review of views concerning the taking of some potential chemopreventive agents. BMC Medical Research Methodology. 2011;11(1):29.