Chapter 9

Making Space for Qualitative Evidence in Global Maternal and Child Health Policymaking

Christopher J. Colvin

Introduction

The success of health interventions often hinges on complex processes of implementation, the impact of sociopolitical and cultural contexts, resource constraints and opportunity costs, and issues of equity and accountability. Qualitative research offers critical insights for understanding these issues. “Qualitative evidence syntheses” (or QES)—modeled on quantitative systematic reviews—have recently emerged as an important vehicle for integrating insights from qualitative evidence into global health policy.

However, it is challenging to integrate QES into policymaking in ways that are both *acceptable* to the often-conservative health policy world and *consonant* with social science’s distinctive methodologies and paradigms. Based on my experiences participating in and observing numerous guideline working group meetings and interviews with key informants, this chapter offers an auto-ethnographic account of an effort to integrate QES into the World Health Organization’s global OptimizeMNH guidelines for task shifting in maternal and newborn health (MNH).

“Global guidelines,” like those developed by the World Health Organization (WHO) and other major international health institutions, are a critical component of global health policy. They synthesize evidence on key policy questions and set norms and standards for health decision-making and practice by “assist[ing] providers, recipients and other stakeholders to make informed decisions about appropriate health interventions” (WHO, 2003). WHO guidelines, in particular, carry significant authority in global public health and have an outsize influence on health decision-making across a wide range of contexts (Ruger & Yach, 2009).

C. J. Colvin (✉)

Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA
e-mail: cjc5r@virginia.edu

As a technology for consolidating the latest knowledge and best practices, and for setting policy norms and standards, guidelines are a recent development, emerging only in the last three decades (Bhaumik, 2017). The birth of the health policy guideline was closely tied to the rise of “evidence-based medicine” (or EBM). EBM has become the hegemonic framework in contemporary biomedicine and health policymaking for legitimizing knowledge claims. At the methodological heart of EBM and “evidence-based policy-making” (EBPM) is the systematic review, which involves the identification, collation, and synthesis of all the available (quantitative) evidence on the safety, efficacy, and cost-effectiveness of health interventions (Mykhalovskiy & Weir, 2004).

Guidelines are typically developed in response to specific policy and practice questions that, in turn, reflect knowledge gaps or contentious debates within the field. Systematic reviews are meant to provide independent, “evidence-based” answers to these questions as well as guidance on policy options and preferences. During the first few decades of EBM—from the mid-1970s to the early 2000s—the preferred form of evidence in systematic reviews was the randomized controlled trial (RCT), a method for producing evidence of the efficacy of a discrete intervention by attempting to control—through randomization—for all other factors related to an outcome of interest (Adams, 2013). Other forms of quantitative evidence that also assessed efficacy, albeit in a less controlled fashion, were included in these reviews but fell further down EBM’s standard “hierarchy of evidence.”

Though the language and logic of EBM and the RCT are ubiquitous in biomedicine and health policy, there is growing recognition of the limits of standard EBM approaches (Ioannidis, 2014; Hutchison & Rogers, 2012) as well as the potential contribution of qualitative evidence (Gilson et al., 2011). Those involved in health policy are increasingly linking the success of health interventions to wider questions about the complex processes of program implementation, the impact of socio-political contexts, resource constraints and opportunity costs, and issues of local agency, equity, and acceptability (Olivier de Sardan et al., 2017). Qualitative research is seen by many to offer a critical evidence base for addressing these questions (Lewin & Glenton, 2018).

In order to integrate qualitative research into EBM’s existing evidentiary practices, some researchers have been developing methods for including qualitative evidence in policymaking in ways that are accepted as legitimate by those working within orthodox EBM frameworks. “Qualitative evidence syntheses” (or QES)—modeled on principles and methods of quantitative systematic reviews—is one such methodology. QES has been practiced on a small scale for a number of years but has only recently gained momentum as a method for integrating qualitative research into more formal processes of health policy and decision making.

The growing place of QES in health policy raises a number of important questions about new (and old) forms of evidence production, circulation, and consumption and about the ways in which these forms of knowledge practice reflect, and in turn have an impact on, forms of responsibility and accountability within health policy, and health governance more broadly. What forms of qualitative evidence will be included (or excluded) in decision making and how will this evidence be

evaluated (and even ranked)? How will global policymakers understand the relationships (and differences) between quantitative and qualitative forms of evidence? How will they integrate knowledge from qualitative research about the role of local context into guidelines often understood as universal? Does the mere presence of qualitative evidence have the power to change global health decision-making processes and politics for the better, or will it simply be co-opted in service to already existing goals and perspectives?

This chapter tackles some of these questions through the lens of an ethnographic case study of the use of QES in the development of the WHO's "OptimizeMNH guidelines" (WHO, 2012) for task shifting in maternal and newborn health (MNH). The Optimize guidelines were the first WHO guidelines to officially include QES in the evidence base that informed its recommendations. It thus provides an excellent opportunity to reflect on the emergence and significance of new forms of knowledge production and decision making in global health policy. It is also an opportunity to reflect on the potential connections and tensions between the field of maternal and reproductive health and these new forms and practices of evidence-into-policy work. The ethnographic focus of the chapter is on the Optimize Technical Working Group (TWG), a collection of WHO staff and external methodologists and researchers who designed the guideline development process and carried out many of the required evidence syntheses. The TWG was a rich ethnographic site for charting the development of new methods for integrating qualitative evidence into existing health policy practices.

In the end, the OptimizeMNH guidelines were perceived by those involved to be a great success. The qualitative evidence syntheses we produced played an important role in shaping the guidelines' recommendations. The TWG enjoyed flexibility in how it approached its work as well as widespread acceptance, even enthusiasm, about its efforts among the broader group of policymakers engaged with the guidelines. The process did require negotiating an often delicate balance between existing quantitative-focused paradigms and methods for evidence synthesis and decision making, and the distinctive paradigms and methods of qualitative research. But the case study also demonstrated that EBM knowledge production practices can be more flexible and more accommodating of multiple methodological and epistemological perspectives than has been previously reported.

A Note on Methods

I first became involved in Optimize not as an ethnographer, but as a direct participant when I was commissioned to work on several of the qualitative evidence syntheses for the guideline and join the TWG. I was initially asked to participate because of my previous research on task shifting among community health workers (CHWs). I had never conducted a QES before nor had any exposure to guideline development. Given the recency of both QES methodology and the evidence review process for Optimize, however, I was told that my lack of experience was not an

issue. As soon as I got involved in the project, I became fascinated by the space for qualitative research that had apparently opened up in this particular domain of health policy. I began developing a parallel (auto-)ethnographic project to reflect on how this space opened up, how it is understood, experienced, and negotiated by those involved, and what it might (or might not) say about changing practices of knowledge production in health policy and practice. The impetus for developing this parallel project has been my ongoing experience of shifting—sometimes uncomfortably—between my anthropological lens on the world, and my public health perspective. The critical medical anthropologist in me was—and remains—suspicious of biomedicine’s newfound interest in “qualitative” work; but the public health researcher in me saw—and still sees—this opening as an opportunity to be met with (cautious) good faith.

I have remained involved in a number of projects of this kind since my experience with Optimize, many of them with the same people in that initial TWG. The material presented here is based on my reflections about participating in the Optimize guideline development process as well as numerous informal conversations over the years and 12 formal in-depth interviews with several of the TWG members and others involved in QES work. My work on these guidelines began in 2011 and I have been working slowly on this parallel ethnographic project since then. My source material comes from field notes, minutes of meetings, formal review and guideline documents, and interview notes. The process of analysis has been iterative and has unfolded over time, through conference presentations and conversations with those working with me in this field. I have continued to discuss this parallel ethnographic project with the TWG group members and they have remained enthusiastic about the chance to reflect on this work. Ethics approval was provided by the University of Cape Town’s Faculty of Health Sciences Human Research Ethics Committee.

A Primer on Qualitative Evidence Synthesis

The recent growth of qualitative evidence synthesis as a method for “secondary” knowledge production emerges from the intersection of two parallel developments in global health research. The first is the aforementioned rise of evidence-based medicine and the complex set of evidentiary practices, knowledge claims, and forms of accountability and audit that have accompanied EBM (Adams, 2016). The second, less visible development has been the steady but quiet growth of “qualitative health research” in the health sciences (Pope & Mays, 2009). Though the tensions between qualitative and quantitative research in health (Inhorn, 1995; Petticrew & Roberts, 2003; Porter, 2006) persist, there has nonetheless been a slowly growing inclusion of primary qualitative research within the health sciences (Shuval et al.,

2011).¹ This inclusion of qualitative research has often been slow, grudging, and conducted on the terms of more powerful actors within health research (see, for example, the recent debate (Daniels et al., 2016; Greenhalgh et al., 2016) around the inclusion of qualitative research in the *British Medical Journal*). Nonetheless, there are significantly more qualitative health researchers, research units, academic journals, and grant-funded research projects than there were at the beginning of EBM's rise.

The focus of this chapter is not, however, on the rise of primary qualitative research in health but rather on the more recent emergence of “qualitative evidence synthesis,” a term for the broad cluster of methods for systematically synthesizing the findings of primary qualitative research from multiple research studies. QES methods follow, in many ways, the basic logic of a quantitative systematic review: a thorough search of the available evidence around a specific review question, an assessment of the quality of each of these underlying studies, clear guidelines for inclusion and systematic procedures for data analysis, and the use of multiple reviewers and audit trails. However, QES approaches tailor their techniques to the important methodological and epistemological differences between quantitative and qualitative research (Hannes & Macaitis, 2012).

There has been rapid growth in QES, with nearly all of the existing reviews—now numbering in the thousands—having been conducted in the last 15 years, the majority of them in the last 5 years. Authors use a wide, sometimes dizzying range of approaches in QES (Dixon-Woods et al., 2005), reflecting the diversity of methodological and epistemological approaches that fall under the broad category of “qualitative research.” Most QES authors work in applied contexts and most of the existing syntheses have been produced by researchers in the UK, parts of Europe, and Australia.

Though QES generally entail a high degree of methodological complexity and systematicity and are often published in peer-reviewed journals, they are frequently initially produced for use by health policymakers and practitioners in relation to specific policy and practice questions. Proponents of QES make several arguments for how these syntheses add an important dimension to health policymaking. They argue that qualitative evidence can answer questions that are distinct from questions asked in quantitative research, including: (1) the scope of health problems (and their causes), (2) the perspectives and experiences of different groups of people in relation to these health problems, (3) the acceptability and feasibility of interventions to address health problems, and (4) factors that affect the implementation of these interventions (Lewin & Glenton, 2018).

¹In evidence synthesis, a distinction is typically made between “primary research” that collects and analyses new data (e.g., from interviews, observations, document reviews, and ethnography) and secondary research which synthesizes the findings from reports of previously conducted research (e.g., from journal articles, books, NGO, or government reports).

Origin Stories: QES in OptimizeMNH

The Optimize guidelines aimed to contribute answers to all four of these types of questions in relation to “task shifting” in maternal and newborn health programs. Task shifting describes the reorganization of discrete health service tasks by moving them from one cadre of health worker to another (WHO, 2006). Though task shifting initiatives can address a variety of problems, in most cases, it is a response to the global crisis in human resources for health and an effort to realign service delivery to be more efficient and more in line with local resource limitations and capacity constraints (Lehmann et al., 2009; Mishra et al., 2015).

In this case, practitioners, program managers, and policy-makers in maternal and newborn health were confronted with a series of unanswered questions about task shifting. The WHO Departmental Director who oversaw the guideline development reported that the initial trigger for Optimize was debates around whether CHWs could safely and effectively administer misoprostol to treat postpartum hemorrhage (PPH) outside of health facilities. Misoprostol is a cheap, easy-to-use, and highly effective life-saving drug that can prevent a major cause of maternal mortality (Smith et al., 2015). Reproductive health activists were pushing for wider use of misoprostol by CHWs for those women who could not or chose not to deliver in health facilities. WHO country offices were also asking for guidance on this issue and there was a long-standing debate within the scientific literature as well. A number of contentious issues were at play, including the medical risks of improper administration of misoprostol, fears that administration by CHWs would disincentivise facility birthing, and the fact that misoprostol can also be used for medical abortion (Wainwright et al., 2016).

Though task shifting of misoprostol administration for PPH was the initial trigger for Optimize, early discussions on the scope of the guideline soon identified a wide range of other task shifting questions in maternal and newborn health around which there was uncertainty about safety, efficacy, cost-effectiveness, acceptability, and feasibility. There were questions, for example, as to whether CHWs could be trained to perform neonatal resuscitation or nurses could be trained to perform emergency caesareans. The goal of Optimize was to review the available evidence on task shifting those key tasks and cadres around which there was uncertainty. The conventional approach in this kind of scenario is for a team like the Technical Working Group to compile the existing systematic reviews of safety and effectiveness, commission systematic reviews in areas where they do not yet exist, and present all the available evidence to an independent panel, the Guideline Development Group (GDG), that would then develop the official guideline recommendations.

In the earliest discussions around Optimize, the focus was indeed on collecting quantitative evidence to address safety and effectiveness. It did not take long, however, for those involved to recognize that task shifting is as much a problem of program implementation and local context as it is one of technical training and clinical service provision. There were also larger programmatic and political issues at stake here, including: the role of community health workers in primary health

care; the tensions between midwives, nurses, and doctors; the debates over facility-based versus home-based delivery; and the generally poor quality of care for women and newborns in many countries. Those involved in these early discussions recognized these were questions that could not be resolved through RCTs; they needed qualitative research to more fully understand these issues and the possible ways to address them.

This recognition of the need for qualitative research did not, however, emerge only in relation to the specific policy questions at stake in Optimize. There were a number of people, institutions, resources, and existing relationships that aligned to set the stage for recognizing the value of qualitative research in this particular set of guidelines. During informal discussions about the project, for example, TWG members often highlighted the positive influence of the Director of the Department producing the guidelines, a medical doctor with no background in qualitative research but with deep programmatic experience of the implementation of MNH services. The TWG's methodological experts had an existing relationship to this Director and were also long-standing advocates for improving how guidelines were developed and expanding the bodies of evidence guidelines made use of.

These more internal dynamics within the WHO were complemented by recent expansions in the broader conceptual vocabularies of global health policymakers and practitioners, including the rise of "health systems thinking," complexity theory, and implementation science, all important contexts for Optimize. In the years preceding the guidelines, there was growing international interest in health systems research and program implementation, and there were several initiatives just at the WHO to improve how guidelines were developed for health systems problems (Bosch-Capblanch et al., 2012). These conceptual developments in health systems research, along with the steadily growing presence of qualitative health research more generally, helped highlight the utility of including qualitative evidence in guidelines like Optimize.

Just as critical, however, was the work of individuals and research and policy institutions—like the Norwegian Knowledge Centre for the Health Services, the UK's National Institute for Health and Care Excellence (NICE), and the Joanna Briggs Institute—that had been working for years at the intersection of qualitative health research and policymaking, developing the concepts and tools, forming the personal and professional relationships, and consolidating the human and financial resources necessary for such an undertaking. Some of this foundation-laying took the form of active lobbying and awareness-raising to promote greater inclusion of qualitative evidence. For the most part, however, it was the alignment of numerous individuals, institutions, and ideas working with or, more often, adjacent to each other over time that ultimately created the window of opportunity for Optimize's engagement with qualitative evidence.

New Hierarchies Undone by Old Methods

The core technical working group that I joined was compromised of eight members. In addition to the Department Director mentioned earlier and me, there were three other anthropologists, two with QES and guideline development experience; a nursing professor with QES methods expertise; and two medical doctors with some recent experience in conducting or reviewing qualitative research. Training in anthropology was predominant in the group and in our conversations, all of the group members (including the clinicians) frequently spoke of ethnography as both the richest and most rigorous and trustworthy form of evidence we might include in our syntheses. Though never formally codified as such, long-term, in-depth ethnographic studies stood at the top of a now-inverted hierarchy of evidence for QES.

Since there were no existing QES that addressed the guideline's focus on task shifting for maternal and newborn health, we decided to conduct three new syntheses, on lay health workers (LHWs), midwives, and mid-level providers (who occupy a space between nurses and doctors in the professional hierarchy). In each review, the primary objective was to understand the factors affecting implementation of task shifting initiatives in maternal and child health programs. These factors included both familiar concepts such as the "acceptability" of such initiatives among health-care workers, patients, family members, and others as well as concepts related to broader social and political contexts that could affect the "feasibility" of these programs.

At the start of the process, I, at least, had the sense that there was a rich ethnographic literature we would be able to draw on for the LHW and midwife reviews. I knew some of this literature already and imagined spending the next few months reading more deeply. In practice, however, rather than diving into ethnographies, I spent most of my time reading quite short, poorly conceived and poorly written qualitative health research articles based on once-off interviews and focus groups with small groups of people. In these articles, the methods section was sometimes as long as the main body of the paper and the most common analytic strategy an anodyne form of "thematic analysis." Quotes peppered the brief text but little was offered in the way of context or interpretation and discussion.

The reason we spent most of our time working through this literature was primarily methodological. The logic of systematic reviews requires a clearly defined research question and a systematic search strategy. This is consonant with the reductionist approach of the natural sciences, where the phenomenon of interest is narrowly delimited and closely examined under controlled circumstances. In our reviews, we defined our questions clearly but also, we thought, quite broadly. The LHW review, for example, was interested in any evidence on any shifting of health service tasks to or from LHWs in the context of maternal and newborn health. In theory, any ethnography that included LHWs working with women and babies would have been eligible.

In practice, however, finding such literature efficiently was next to impossible. First, few ethnographic studies define themselves narrowly around specific groups

such as LHWs or specific concepts such as task shifting. Task shifting might play a key role in an ethnographic text but unless it is indexed as such in a database, it would be very difficult to find efficiently. Ethnographic research rarely sets out to answer such narrowly defined research questions using such discrete variables, and rarely frames its questions in the vocabulary of the health sciences and health policy.

Second, we had decided to not search for books, though we could include them if we knew of them already or found them during our searches and they met our inclusion criteria. This was primarily because most databases that catalogue books do not do so to the same degree of precision as journal articles. The kinds of searches that are standard practice in EBM—precise and multi-layered searches of abstracts, keywords, full-text, forward and backward citations, etc.—are generally not possible with books. The inclusion of books would have also represented a pragmatic dilemma in terms of how long it might have taken to properly read and extract the findings from multiple books.

Finally, ethnographic and other social science research articles published outside of biomedical and public health contexts could be difficult to find because of the poor quality of the database search engines that indexed these publications. The biomedical research world has, over many years, and as a result of the rise of EBM, slowly but significantly improved the platforms for indexing and searching this literature. Whereas this literature is often richly “tagged” with meta-data and is searchable with complex algorithms, the important anthropological journal database AnthroSource at the time only offered keyword searching, and frequently missed relevant articles we knew existed within its publications. Some medical anthropology journals are indexed and searchable in PubMed but this was only a limited solution to a deeper problem.

As a result, the literature that we had imagined sitting atop the QES hierarchy of evidence was precisely the literature that was least visible to the methods and tools we had at hand. Furthermore, the literature we did find from anthropology and related disciplines was most often focused on the daily lived experience and “cultural contexts” of individuals, communities, and healthcare workers. As valuable as this was, it did not meet our inclusion criteria. Ethnographic work that closely examined the process of the implementation of task shifting interventions was much harder to come by.

Throughout the search process, we thus confronted a number of obstacles in our ambition to mine the ethnographic literature. These barriers included the form and content of ethnographic knowledge itself, as well as the ways in which this knowledge is archived and made visible and accessible to other researchers. Also critical was the limiting role of time and financial constraints in determining what kinds of searches were possible and what kinds of knowledge would be feasible to include in the review.

NerdWorld: Pragmatism, Innovation, and Ideology

I was at first frustrated by the methodological barriers to including ethnographic literature. These barriers were not merely technical but rather the result of long-standing techniques for ordering knowledge and privileging certain kinds of research questions and methods. It was also clear, however, that there was no easy way around this dilemma, not without a wholesale restructuring of the logic and methods of systematic reviews. This was something that would have not only required significant time and resources but would have also scuttled the effort to integrate QES into these guidelines.

Despite this particular methodological impasse, though, on the whole, I found the TWG's approach to methods to be pragmatic and flexible. When I started, I feared two types of methodological orthodoxy. First, I was concerned that the group members might try to mechanically and inappropriately translate techniques for quantitative systematic reviews to the qualitative literature. It was clear to me how this approach would undercut any legitimate integration of qualitative research into the guidelines.

Second, I was worried about another kind of orthodoxy, dominant in some of the qualitative health research methods literature. This part of the literature is characterized by a rigid, defensive, and even mystifying approach to qualitative research methods that tries to define itself in distinct contrast to quantitative methods. It foregrounds presumably insurmountable differences in epistemology between qualitative and quantitative paradigms and advertises the complex, thorough, and rigorous nature of qualitative data analysis (often through highly formalized procedures and complex terminology). This approach diverges dramatically from my methods training in anthropology, which like many of my generation primarily consisted of advice to just go to the field and "figure it out."

However, the TWG's methodological experts generally took a pragmatic approach. When we asked how we could adapt our search strategies, or choose a conceptual framework for analysis, or determine the difference between "empirical" research and "opinion" pieces, or draw inferences from "indirect" evidence on different but closely related topics, the methodologists generally simply asked the group what seemed most "reasonable." These kinds of questions were often debated at length in group discussions, where all sides of the issues were put on the table and a feasible but reasonable course of action eventually set out. We were always reminded at the end of these discussions to "show our work" and explain our reasoning.

Quantitative research is of course also filled with judgment and pragmatic choices like these, if mostly implicitly. It is no less shot through with the need to manage ambiguity, uncertainty, incompleteness, and the hard-to-predict impact of context. But there are well-established procedures in quantitative research—again, many of them implicit—for making this messiness much less visible. What felt different here was the absence of widely recognized, pre-existing standards against which to measure (or obscure) our pragmatic choices. In their absence, we were

counseled to record and justify our approach. This imperative to “be transparent” was intended to both ensure accountability and lay the foundation for the next round of researchers to build upon and develop a community of practice for these new methodologies.

There was a great deal of positive energy and engagement in these methodological discussions, as well as an explicit contrast drawn between the routine pragmatics of methodology and broader ideological battles around knowledge production. One TWG member described the group as a bunch of “nerds” more interested in crafting a “good enough” methodology than in litigating ideological debates about qualitative versus quantitative epistemology. These methodological discussions were also characterized by an excitement around the chance to develop new methods for new problems and to be involved in the “ground-breaking” step of including QES in WHO guidelines.

The work of the TWG and the perspectives of its members were, of course, shot through with various ideologies—not least of which was the focus on innovation and the “historic” nature of the group’s work—but it was not always clear which ideological frameworks were at play at a given time. Certainly, one ideological feature that often sat in tension with our pragmatic impulses was the emphasis on “evidence-based methods” and the need for vigilance in preventing unwarranted and “subjective” assumptions about methodological choices to overrule evidence-based decisions. For example, a debate emerged during the review design process about how many studies to include. Rather than aim for an exhaustive review of all the available evidence that is possibly relevant (as in a quantitative review), we did what social scientists often do—searched for and included the best evidence that would tell us the most, however, that was defined—about what we were interested in. We did this by sampling purposively from the broad set of studies identified in our original search.

This rather conventional approach generated some real hesitation among some team members about the potential for bias and the lack of an objective evidence base about how certain sampling strategies might shape the direction of our review. The anthropologists in the group countered that the possibility of alternative interpretations was a central tenet of the interpretive process—and thus not a threat, nor something that could be resolved through “evidence”—and that a random sample of eligible studies would send precisely the wrong message about how interpretation works. We also argued that reading the full set of potentially eligible studies, in some effort to avoid “bias,” would likely overwhelm our capacity to effectively interpret it. Whereas “more data” is often assumed, rightly or wrongly, to be preferable in quantitative research, this is usually not the case in qualitative research.

Behind this decision to purposively sample were other factors as well, including limited funding, limited time frames, and a palpable dread within the TWG of having to deal with even more poorly done qualitative health research. These factors—time, resources, and the often-difficult lived experience of the review process—were significant forces shaping the pragmatic decisions made within our reviews.

Show Your Work!: Transparency, Accountability, and Interpretation

While this kind of pragmatism was a welcome surprise to me, one consequence of “showing your work” was a kind of “bread crumb” approach to supporting arguments and referencing studies in our narrative. See, for example, the following excerpt from the LHW review:

“In four USA-based studies (S16; S50; S52; S53), LHWs gave mothers, including teenage mothers and others in difficult socioeconomic circumstances, emotional and practical support and also promoted healthy behaviours during pregnancy, childbirth and in the first few weeks after birth. In one Australian study, LHWs offered emotional support and practical help to parents at risk of child abuse and neglect (S55). In approximately twelve studies, from Australia (S10), Canada (S17-18, 48), the UK (S35, S37, S44), the USA (S40, S46), Brazil (S47), Mexico (S51), India (S1), Papua New Guinea (S2) and Viet Nam (S19), LHWs carried out a package of tasks that were primarily promotional with the aim of improving maternal and child health” (Glenton et al., 2013).

This kind of analysis is at the opposite end of the spectrum from my experience in anthropology where ethnographic authority (properly established), the ethnographic vignette, and convincing theoretical argumentation are valued as the proper grounds for defending interpretations. Ethnographic claims should be defensible but not entirely or, at least, easily auditable.

The approach here, however, is organized in a way that always anticipates the audit. There was even frequent mention of the need to leave “audit trails” for future reviewers and readers. In the review narrative, however, this rhetorical strategy sits in awkward tension with conventions in qualitative interpretation. This is perhaps best seen in the strange phrase “In approximately twelve studies,” a nonsensical wording (since it is *exactly* 12 studies) that attempts to combine the imperative for transparency and audit with qualitative narrative’s customary resistance to quantifying individual points of information.

This approach is rooted in an idea that one can directly link interpreted findings to individual findings in the individual studies being reviewed. Instead of building holistic interpretations within studies or developing thematic, even explanatory interpretations across studies, the primary unit of analysis was the individual finding, supported by a precisely specified evidence base.

In the process, however, the cited studies appear to provide equivalent support for each claim. Each study, though, was conducted in a particular context with its own unique research question. The collected findings may have only been loosely comparable to each other. Specifying all the possible nuanced differences between studies supporting a synthesized finding is impractical (and is in fact the necessary “hidden” work of interpretation). But the demand for transparency required this kind of documenting and it often felt like the text—and our analytical ambition—was weighed down by these citations, making it difficult to draw connections across findings, evaluate patterns, and propose interpretations at a more abstract level.

Much of this effort to show our work and be transparent in our methodological judgments and interpretations was about demonstrating the legitimacy of this type of evidence production and our willingness to be held accountable for our work. This was especially important given the skepticism with which we feared some in the WHO and elsewhere would view our efforts to integrate qualitative evidence into these guidelines. Other forms of accountability and legitimacy, both new and old, were at work here as well, though. Central to the perceived legitimacy of our efforts was that we were working under the auspices of the WHO's overall Guidelines Review Committee and its *Guidelines Development Handbook*. Official institutional approval, and our accountability to the institution's procedures and standards, was vital to the legitimacy of our work, both in the eyes of the TWG members and other observers. Also important was the fact that many TWG members were affiliated with recognized research institutions and initiatives, ensuring a further layer of legitimacy and accountability.

There was one final and unexpected dynamic of accountability that may have emerged out of our careful documenting of the qualitative evidence. When this project began, I heard concerns from several people about the ways Guideline Development Groups (GDG)—the independent experts who review the TWG's work and make the formal recommendations—sometimes too easily brought their own personal experiences and perspectives to the table, using their social status within their field or the GDG to push specific policy options. This was especially true, they argued, when a GDG had only had effectiveness and safety data to work with. There have also been long-standing concerns that the experiences, needs, and preferences of patients and affected communities are not adequately represented in GDGs and that specialist panelists may too often think they can “speak for” these under-represented groups (Knaapen & Lehoux, 2016).

We certainly did not expect our qualitative syntheses to provide data-driven answers to all of the possible policy questions and options on the table. Having the evidence from our reviews readily at hand, however, may have transformed to some extent how conversations and decisions unfolded within the panel. There may have been less scope for bringing personal experiences with implementation of programs to the conversation when evidence on implementation was now available in another form. The presence of qualitative reviews may have also sat in tension with the inputs from stakeholder, advocacy, and community representatives on the GDG.

Conclusion

So what might the aforementioned developments signify? I have described how those of us working to integrate QES in the Optimize guideline development process attempted to open up a space for a new kind of evidentiary practice—one which could in turn entail new forms of policy decision-making and practice—but were then confronted by a range of social, material, and technical limits to this project. Observers and supporters of this effort spoke often of the rhetorical importance of

having QES meaningfully integrated into these guidelines, arguing that it would strike a real blow for the inclusion of new and potentially transgressive forms of knowledge within the often conservative world of health policy. I have mentioned earlier the growing strength of ideas of health systems, implementation science, and policy translation in global health as well as the slow (and still siloed) growth of qualitative research in health (Daniels et al., 2016). The decision to integrate QES into these WHO guidelines was indeed a coup for those promoting these agendas.

Of course, in practice, the project's success hinged on a careful balance between adopting conventional, quantitatively inspired methods and standards familiar to the WHO authorities overseeing the guidelines and innovating new methods that would both suit the project needs and please a different set of authorities (i.e., social scientists and other qualitative research experts). This experience raises important questions about how our understandings of EBM and EBPM may need to be revised.

For example, the integration of QES in global guideline development entails many of the same processes of *formalization* (of synthesis methods), *disentanglement* (of evidence from its underlying contexts), and *separation* (of science from politics) that have been identified by ethnographers and STS scholars of EBM (Moreira, 2007; Bohlin, 2012; Sundqvist et al., 2015). However, many of these procedures also look somewhat different from how they have been described in the literature on quantitative-focused EBM. The use of purposive sampling in QES, for example, introduces forms of judgment typically thought—albeit incorrectly—to be excluded from EBM procedures. Similarly, recognition of the critical role of local context in explaining the findings from individual studies sits in tension with the idea that EBM always works to disentangle findings from the contexts of their production. Finally, the broad scope and flexible nature of many QES review questions and the often iterative nature of QES search processes both complicate the production of “non-evidence” in EBM (Knaapen, 2013). While QES do require decisions about what should be included as “evidence” that is responsive to a particular review question, the judgments required to decide on the boundary between evidence and non-evidence look different (see Noyes (2018) for a discussion of the question of relevance when assessing evidence included in QES).

There are signs of the uptake of QES approaches in guidelines across a wide range of WHO departments (and at other institutions such as the UK's National Institute for Health and Care Excellence (NICE)). The Department Director who initiated the Optimize project identified some significant ongoing effects in his field of the integration of QES in WHO guidelines. He described a recent guideline his department produced on antenatal care (ANC) (WHO, 2016) that normally would have focused on biologically framed outcomes like ANC-related morbidity and mortality. After his department's experience with Optimize, however, the scope of the guidelines was expanded to include “positive pregnancy experiences,” a phrase they even included in the guideline title and as a core guideline objective. As with Optimize, concerns with biological outcomes were integrated with social, cultural, psychological, and experiential outcomes when developing recommendations.

Such an extension of the imaginative boundaries of one or two WHO guidelines is not revolutionary. It is perhaps better understood as the result of long-term

foundation-laying work in the fields of health policy and systems research—and qualitative health research more generally—as well as growing recognition of implementation processes and local contexts as critical factors in the success of any health intervention. At the very least, the growth of QES in global health policy signals that there remains room—perhaps growing room—in some spaces for a more expansive vision of what is at stake in health policy and what forms of knowledge might best contribute to decision making about health.

References

- Adams, V. (2013). Evidence-based global public health: Subjects, profits, erasures. In J. Biehler & A. Petryna (Eds.), *When people come first: Critical studies in global health* (pp. 54–90). Princeton University Press.
- Adams, V. (2016). *Metrics: What counts in global health*. Duke University Press.
- Bhaumik, S. (2017). Use of evidence for clinical practice guideline development. *Tropical Parasitology*, 7(2), 65–71.
- Bohlin, I. (2012). Formalizing syntheses of medical knowledge: The rise of meta-analysis and systematic reviews. *Perspectives on Science*, 20(3), 273–309.
- Bosch-Capblanch, X., Lavis, J. N., Lewin, S., Atun, R., Rottingen, J. A., Droschel, D., et al. (2012). Guidance for evidence-informed policies about health systems: Rationale for and challenges of guidance development. *PLoS Medicine*, 9(3), e1001185.
- Daniels, K., Loewenson, R., George, A., Howard, N., Koleva, G., Lewin, S., et al. (2016). Fair publication of qualitative research in health systems: A call by health policy and systems researchers. *International Journal of Equity in Health*, 15, 98.
- Dixon-Woods, M., Agarwal, S., Jones, D., Young, B., & Sutton, A. (2005). Synthesising qualitative and quantitative evidence: A review of methods. *Journal of Health Services Research and Policy*, 10(1), 45–53.
- Gilson, L., Hanson, K., Sheikh, K., Agyepong, I. A., Ssengooba, F., & Bennett, S. (2011). Building the field of health policy and systems research: Social science matters. *PLoS Medicine*, 8(8), e1001079.
- Glenton, C., Colvin, C. J., Carlsen, B., Swartz, A., Lewin, S., Noyes, J., et al. (2013). Barriers and facilitators to the implementation of lay health worker programmes to improve access to maternal and child health: Qualitative evidence synthesis. *Cochrane Database of Systematic Reviews*, 10, CD010414.
- Greenhalgh, T., Annandale, E., Ashcroft, R., Barlow, J., Black, N., Bleakley, A., et al. (2016). An open letter to the BMJ editors on qualitative research. *BMJ*, 352, i563.
- Hannes, K., & Macaitis, K. (2012). A move to more systematic and transparent approaches in qualitative evidence synthesis: Update on a review of published papers. *Qualitative Research*, 12(4), 402–442.
- Hutchison, K. J., & Rogers, W. A. (2012). Challenging the epistemological foundations of EBM: What kind of knowledge does clinical practice require? *Journal of Evaluation in Clinical Practice*, 18(5), 984–991.
- Inhorn, M. C. (1995). Medical anthropology and epidemiology: Divergences or convergences? *Social Science and Medicine*, 40(3), 285–290.
- Ioannidis, J. P. (2014). How to make more published research true. *PLoS Medicine*, 11(10), e1001747.
- Knaapen, L. (2013). Being ‘evidence-based’ in the absence of evidence: The management of non-evidence in guideline development. *Social Studies of Science*, 43(5), 681–706.

- Knaapen, L., & Lehoux, P. (2016). Three conceptual models of patient and public involvement in standard-setting: From abstract principles to complex practice. *Science as Culture*, 25(2), 239–263.
- Lehmann, U., Van Damme, W., Barten, F., & Sanders, D. (2009). Task shifting: The answer to the human resources crisis in Africa? *Human Resources for Health*, 7, 49.
- Lewin, S., & Glenton, C. (2018). Are we entering a new era for qualitative research? Using qualitative evidence to support guidance and guideline development by the World Health Organization. *International Journal of Equity in Health*, 17(1), 126.
- Mishra, S. R., Neupane, D., Preen, D., Kallestrup, P., & Perry, H. B. (2015). Mitigation of non-communicable diseases in developing countries with community health workers. *Globalization and Health*, 11, 43.
- Moreira, T. (2007). Entangled evidence: Knowledge making in systematic reviews in healthcare. *Sociology of Health and Illness*, 29(2), 180–197.
- Mykhalovskiy, E., & Weir, L. (2004). The problem of evidence-based medicine: Directions for social science. *Social Science and Medicine*, 59(5), 1059–1069.
- Noyes, J., Booth, A., Lewin, S., Carlsen, B., Glenton, C., Colvin, C. J., et al. (2018). Applying GRADE-CERQual to qualitative evidence synthesis findings-paper 6: How to assess relevance of the data. *Implementation Science*, 13(Suppl 1), 51–61.
- Olivier de Sardan, J. P., Diarra, A., & Moha, M. (2017). Travelling models and the challenge of pragmatic contexts and practical norms: The case of maternal health. *Health Research Policy and Systems*, 15(Suppl 1), 60.
- Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: Horses for courses. *Journal of Epidemiology and Community Health*, 57(7), 527–529.
- Pope, C., & Mays, N. (2009). Critical reflections on the rise of qualitative research. *British Medical Journal*, 339, 737–739.
- Porter, J. D. (2006). Epidemiological reflections of the contribution of anthropology to public health policy and practice. *Journal of Biosocial Science*, 38(1), 133–144.
- Ruger, J. P., & Yach, D. (2009). The global role of the World Health Organization. *Global Health Governance*, 2(2), 1–11.
- Shuval, K., Harker, K., Roudsari, B., Groce, N. E., Mills, B., Siddiqi, Z., et al. (2011). Is qualitative research second class science? A quantitative longitudinal examination of qualitative research in medical journals. *PLoS One*, 6(2), e16937.
- Smith, H. J., Colvin, C. J., Richards, E., Roberson, J., Sharma, G., Thapa, K., et al. (2015). Programmes for advance distribution of misoprostol to prevent post-partum haemorrhage: A rapid literature review of factors affecting implementation. *Health Policy and Planning*, 31(1), 102–113.
- Sundqvist, G., Bohlin, I., Hermansen, E. A., & Yearley, S. (2015). Formalization and separation: A systematic basis for interpreting approaches to summarizing science for climate policy. *Social Studies of Science*, 45(3), 416–440.
- Wainwright, M., Colvin, C. J., Swartz, A., & Leon, N. (2016). Self-management of medical abortion: A qualitative evidence synthesis. *Reproductive Health Matters*, 24(47), 155–167.
- WHO. (2003). *Guidelines for WHO guidelines*. <https://apps.who.int/iris/handle/10665/68925>. Accessed 1 Feb 2021.
- WHO. (2006). *Task shifting to tackle health worker shortages*. https://www.who.int/healthsystems/task_shifting/TTR_tackle.pdf?ua=1. Accessed 1 Feb 2021.
- WHO. (2012). *Optimizing health worker roles to improve access to key maternal and newborn health interventions through task shifting*. <https://apps.who.int/iris/handle/10665/77764>. Accessed 1 Feb 2021.
- WHO. (2016). *WHO recommendations on antenatal care for a positive pregnancy experience*. <https://www.who.int/publications/i/item/9789241549912>. Accessed 1 Feb 2021.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.