

Abstract

In the large international projects where many qualitative researchers work, generating qualitative big data, data sharing represents the status quo. This is rarely acknowledged, even though the ethical implications are considerable and span both process and product. I argue that big-team qualitative researchers can strengthen claims to rigor in analysis (the product) by drawing on a growing body of knowledge about how to do credible secondary analysis. Since this necessitates a full account of how the research and the analysis are done (the process), I consider the structural disincentives for providing these. Debates around credibility and rigor are not new to qualitative research in international development, but they intensify when new actors such as program evaluators and quantitative researchers use qualitative methods on a large scale. In this context, I look at the utility of guidelines used by these actors to ensure the quality of qualitative research. I ask whether these offer pragmatic suggestions to improve its quality, recognizing the common and hierarchized separation between the generation and interpretation of data, or conversely, whether they set impossible standards and fail to recognize the differences between and respective strengths of qualitative and quantitative research.

1. 4-5 Key terms: Secondary qualitative data analysis, computer assisted data analysis, ethics, quality, rigor. **Introduction**

The starting point of this special issue is that the emergence and spread of big-data technologies are influencing how small-data qualitative research are being conducted and evaluated (for example, the way computer software is proposed as a replacement for the interpretive skill of qualitative researchers). Within international development, policy makers and research funders place increasing weight on large-scale qualitative research, since rigor is seen as synonymous with larger sample sizes. Qualitative research can be done at multiple scales without losing its ability to address experiences, understandings, and contexts. However, this requires careful work to avoid stripping data of its context due to aggregating analysis across multiple studies and sites, or the requirement to archive data as open access.

In previous work, I have argued that rigor¹ is an ethical issue (Camfield et al., 2014, Camfield & Palmer-Jones, 2015). Researchers have obligations to produce and disseminate high quality, rigorous, robust, and respectful materials. This recognition of the ethical dimensions of rigor is increasingly common among development researchers and evaluators, even where their understandings of what rigor consists in are very different. Compare, for example, Gujit and Roche (2014) who argue that due to the inextricable link between rigor and relevance in evaluation it is best attained using qualitative and participatory methods, and Palermo (2017) who argues that rigor is the exclusive property of Randomized Controlled Trials. Connecting ethics and rigor, however defined, acknowledges that researchers are at least partially responsible for how their data and interpretations are used. Moreover, in a context where resources are constrained and competed over, a single piece of evidence can have considerable weight.

This paper addresses the challenge of achieving rigorous qualitative research within development (see also Johnson & Rasulova, 2017). This challenge is particularly acute in relation to practices of analysis and interpretation, which are rarely described. It occurs in a context of increasingly large-scale qualitative studies, where, as discussed by Hossain and Scott-Villiers, this volume, the person who collects the data is not always the person who analyses it (Camfield, 2014). I discuss the problems this might pose and attempts made to address them through the careful guidance provided to secondary analysts (e.g. Irwin & Winterton, 2011). This addresses the challenge of recreating the context of data generation for analysts who were not involved in the original fieldwork.

An additional concern in relation to rigor, particularly acute in development, is the evidential basis for claims, including how these are warranted (Manski, 2011). Hammersley (2010) makes a useful distinction between data, which is both “given” and “constructed”², and evidence, which is constructed through a process of selection and inference. This process of selection and inference (analysis) is rarely transparent, as I discuss in section 3. The challenge of making it so is described by Greckhamer and Cilesiz (2014, p.425) in relation to discourse analysis where they identify the need to be systematic, searching, transparent, to “warrant with appropriate evidence the study's rigorous

¹ Although simply defined by the Oxford Living English Dictionary (in Jones & Rasulova, 2017, p. 267) as “the quality of being extremely thorough and careful”, within development it usually denotes quantitative and often experimental research and is rarely used to refer to qualitative research. Many qualitative researchers prefer the more open-ended term “trustworthiness” (Guba & Lincoln, 1985), which encompasses credibility, confirmability, dependability, transferability, and authenticity.

² i.e. shaped by the purpose and interests of the collector – see Bakewell (2007).

and systematic analysis process as well as its knowledge claims” and, most importantly, do all this while “producing sufficiently succinct manuscripts.”³

Due in part to lack of clarity over the choices made during analysis, I argue that it is not always clear how strong claims drawn from qualitative research are evidenced. Ethnography in particular faced a crisis of representation in the 1980s and 90s (e.g. Clifford & Marcus, 1986) when critics argued that it depended too heavily on personal experience and embodiment as the source of its authority. One outcome of the ensuing debate was a welcome acknowledgement that any analysis of human systems inevitably involves personal experience and that true objectivity is impossible. This necessitated considerable attention to the stance from which an analysis is made (i.e. the qualities and positions of the analyst). These conclusions on the basis for interpretive authority are relevant to the discussion on secondary analysis in section 2 since it made possible practices such as revisits by later anthropologists (e.g. Freeman, 1983, returning to Margaret Mead’s Samoan field site). These looked at how the original researcher’s conclusion might have been shaped by their positionality and particular encounters/relationships.

Quantitative researchers have faced a similar crisis of representation across the same time-period. Learner (1983, p.37 in Angrist & Pischke, 2010, p.3) famously remarked “hardly anyone takes data analysis seriously. Or perhaps more accurately, hardly anyone takes anyone else’s data analysis seriously.” The lack of transparency around how conclusions were reached, for example, the data transformations that were required, or the analytical “torturing” of the data until it produced significant results has led to a growing practice of replication within economics (Duvendack et al., 2017). This has the same potential as revisits to increase our understanding of and trust in analyses by testing whether these can be replicated (repeated) by an independent observer. Although qualitative data is rarely replicable, due to changing conditions in the field and the inevitable influence of the researcher on the analysis, it is possible to show the effects of different conceptual framings on the results (Savage, 2010). This shows the operational value of Hammersley’s (2010) distinction between data and evidence.

Alongside a broader cultural distrust of experts, evidenced by political events such as Brexit in the UK and the election of Trump in the US, there is increasing distrust in the practices of researchers (Eriksen & Schober, 2017). As I noted earlier, within applied work there is a lot at stake, so why should research users trust blindly? As researchers, should we feel challenged by this, or flattered that our work is being taken so seriously? Corti (2000, p.28 in Mauthner & Parry, 2009, p. 298)

³ One reviewer observed that the requirement for succinctness is the biggest obstacle to providing the level of detail recommended by quality assurance guidelines.

argues that, “if we are to accept the label "scientist", then we should adopt the scientific model of opening up our data to scrutiny, and the testing of reliability and validity.” Although reliability and validity are not usually used in relation to qualitative research, there are qualitative equivalents of equal value developed from Guba and Lincoln (1985) onwards. This challenges Corti’s (2000, section 6.2) claim that “in the UK there are no quality control standards for qualitative research.”

In later work, Corti & Fielding (2016, p. 11) note the “scope that new digital tools provide for ... being able to directly examine the data that a researcher adduces in support of their analysis.” This is part of the potential of big data, which is the focus of this volume. However, it also raises ontological questions around what constitutes data and what is context. In relation to this, they observe the need for guidance for qualitative researchers on what constitutes “supporting data” for an article (ibid:11). Although the requirement to provide supporting data is common in quantitative journals, it has not previously been requested for qualitative, and qualitative and quantitative analysts may have different definitions (for example, whether context is separable from data, as discussed in section 2).

I present the argument of this paper in three parts. First, I briefly review the epistemological, practical, and political debates around secondary analysis, focusing particularly on context and documentation (section 2.) I also highlight a growing body of knowledge about how to do secondary analysis in a credible way. I then look at the potential for computer-assisted qualitative data analysis software (CAQDAS), which is growing in importance in the age of big data, to make analyses more transparent. Despite its promise, I discuss problems that might arise through this process (section 3). Finally, I look at the standards for qualitative research that have developed in different fields since Corti (2000) and assess their utility in assuring the quality of large-scale qualitative research within international development (section 4 and online appendix).

2. Epistemological, practical, and political debates around secondary analysis

The premise of secondary data analysis is that claims to rigor are not rooted in the personal – “I was there so I know” – but can be systematically evidenced through providing a full account of how the research and analysis were done and through access to the original data. Neither of these things are easy to do. Furthermore, the full costs need to be included in funding proposals since it is expensive to address concerns about confidentiality and the potential sensitivity of the data by, for example, reading the complete dataset and removing any information that might inadvertently reveal the identity of the respondent. There may also be structural disincentives for providing a full account of how the research was done. With colleagues, I reviewed the grants funded by the Department for

International Development (DFID) and Economic and Social Research Council (ESRC)⁴ Joint Fund for Poverty Alleviation Research funding stream from 2005 to 2014 and found few researchers were willing to discuss the challenges they had faced, even as part of confidential reporting to the funder (Camfield, in press, Camfield et al., 2016). This makes it much more difficult for secondary data analysts.

Aside from these practical concerns, there has been an epistemological debate over the value of secondary analysis between researchers who work within repositories or projects focused on creating resources for secondary analysis (e.g. Louise Corti, Libby Bishop, Sarah Irwin and Mandy Winterton) and those who feel that secondary analysis involves seeing data in a very particular way (e.g. Natasha Mauthner, Martin Hammersley). Irwin and Winterton (2012, section 2.5) argue, for example, that “primary analysts have a privileged relationship to the data they have generated, but do not necessarily have a privileged claim on the arguments which can be made from that data... ‘being there’ is not the final arbiter of the adequacy of such understandings.” The distinction they make relates to Hammersley’s (2010) distinction between data and evidence. Although there is a temporal and conceptual overlap between data and evidence (the analyzed data provides the grounds for inference and for descriptive and explanatory claims), evidence can be seen as more constructed than raw data. Irwin and Winterton argue that, although primary access to context (what Ottenberg (1990) calls “head notes”) may give researchers a privileged relationship to “data as given,” it doesn’t imply a privileged relationship to “data as evidence” or interpretation (see also Irwin & Winterton, 2011). In fact, Irwin and Winterton (2012, section 2.6) subsequently argue that “there may be a risk that undue emphasis on context and ‘presence’ itself encourages a particular mode of engagement with data, a mode which tends towards description, and a sense that the ‘answers’ reside within the specific data set.” This concern – that proximity generates unhelpfully descriptive analyses – is captured in Daly et al.’s (2007, p. 43) evidence hierarchy, which concurs that descriptive studies are less valuable in informing practice as they “provide helpful lists of quotations but do not offer detailed analysis.” Although Irwin and Winterton (2012, section 5.1) also argue that “overplaying the significance of proximate context relative to other salient factors may risk privileging description over explanation,” from an applied perspective, one needs deep, contextual description in order to understand and change practices in a particular location.

Although advocates of secondary data analysis foreground an ethics of transparency, others argue that there are more profound ethical concerns in imposing a positivist or foundational⁵ model of

⁴ These are respectively the main funders of international development and social science research in the UK.

⁵ One that acknowledges only one social reality that can be captured as data.

data on a very different way of working. Mauthner and Parry (2009, p.294-5) note that, although many see the argument for secondary analysis as successfully concluded through “the inclusion and analysis of the 'contexts' of data generation, preservation and reuse,” these “theoretical and practical 'solutions' neither allay the concerns of many, nor address what are more deep-rooted philosophical tensions.” The reason for this is that they treat context as separate from rather than a constituent part of data. Concerns over the treatment of context are shared by other authors in this volume, recognizing that studies can become separated from their context through the process of sharing data. But given that data sharing is an intrinsic part of most large-scale studies, within the team, if not outside it, my concern in sections 2.1 and 2.2 is to think through how context can be restored to data. The effects of not doing this are a reduction in the accuracy of the data, an increase in the likelihood that it will be misinterpreted, and the flattening out of local concerns and priorities when framed within global analytical frameworks (Hossain and Scott-Villiers, this volume).

2.1 Why does context matter in data analysis?

The debate over the value of experiential vs. recovered context, and even whether context is necessary rather than a potential source of bias, mirrors age-old debates over the subjective nature of qualitative research. The proposal that distance from the data is epistemologically advantageous is a controversial one since, in international work in particular, much of the original meaning may have been lost through processes of translation (Temple, 2005). Unless fieldworkers are exceptionally skilled in note taking or writing reflective accounts of their fieldwork, the implicit context and the embodied experience of the encounter will not be accessible to the analyst. Without this, it seems unlikely that secondary analysts will be able to better understand the meaning of the data than primary researchers, merely by adding meta-data and meta-documentation about the research process or recovering context through field notes, diaries, memos, and talking to the original researchers. Nonetheless, some researchers have taken this position. Fielding (2004, p. 100) argues provocatively that in qualitative evaluation research “to overcome affinities developed in the field, the fieldworker hands over the data to a second team member who will carry out the analysis.” Implicitly, what they lose in context, they gain in impartiality. Copestake et al. (2016) make a similar argument in their design of the QUIP qualitative impact evaluation protocol. They argue that through a process of double blinding (neither the respondents nor the researcher knows which organization the evaluation is being carried out on behalf of) and data analysis by independent analysts in another country, bias can be reduced to a level that would be acceptable in a quantitative evaluation. The need to compare qualitative to quantitative evaluation, and to justify the preference for qualitative using the language of quantitative, reflects the way development actors view qualitative research.

My position is probably closer to Mauthner and Parry (2009, p.299) in arguing that, “the richer the overall data and context are, the better the resulting analyses will be.” It is never possible to wholly grasp the context of a study and the more ambitious studies become, the harder it becomes to be present at every point. Nonetheless, it is important to understand how the contexts of data production, preservation, and analysis are reflected in the data and shape the interpretations it is possible to make from them. The challenge for someone who works in large international teams, but also someone who uses their own data as a teaching tool, is *how* to create, as far as is possible, the experience of being there. This mirrors the challenge in writing a paper or monograph, which is to make the reader feel as though they are alongside you in data generation and interpretation. Faced with the same evidence, they would have made the same judgements. Corti and Fielding (2016, p.4) observe that, “the further we move from direct engagement in data collection activities the further we step away from context.” In these situations, researchers become dependent on others, which, “requires a level of trust and an ability to recreate context.” In the next section I will explore the extent to which context can be recreated.

2.2. How can context be recreated?

As a lead researcher, I take practical steps to ensure that context is preserved. I train researchers in note taking and provide field diaries for them to record observations and reflections. I provide questions for structured reflection to accompany transcribed data. I audio-record all formal encounters and make use of photos and other visual elements to add texture to the data. I organize and record daily debriefings and longer, more reflective debriefing after the end of each site. Finally, I take every opportunity to accompany teams to the field during training and piloting. When I share data with students, I try to provide as much information about the project, the fieldwork, and the sample as possible. This includes “documentation about the project’s origination and fieldwork approach, a topic guide covering questions asked, and metadata about each interview, such as respondent characteristics and interview settings” (Corti & Fielding, 2016, p. 4). It might also include what Corti and Fielding call “rich description” (I often ask researchers to write a detailed, first-person account of their entry into a site), databases in qualitative data analysis packages where researchers will have coded data and recorded memos reflecting on this process, and project notes, for example, on the fieldwork situation. There is scope for including even more relevant context than this. Wilson (2014, section 5.2) argues that analysts need “information on the type of tenure in which respondents live, the schools attended, parental occupations and changes in these over time... [and] pen portraits of respondents with detailed cross-interview tabulations of their socio-economic circumstances.” This will enable them to understand the extent to which they can generalize from their sample by locating it against the population it was drawn from and understanding the ways in

which it differs. An exceptionally thorough set of documentation could also include in-depth interviews with researchers who have played different roles within the project, a bibliography of the work that influenced project development and the disciplinary and research histories of the original research team (Corti & Fielding, 2016; Wilson, 2014).

Although projects that are funded to act as models for others considering archiving may go to these lengths, my experience in the DFID-ESRC analysis mentioned earlier is that the majority do not. Data archiving was a requirement of ESRC funding and 51% of studies expressed their intention to archive. Nonetheless, only 16 out of 75 completed studies had archived data. Of those, two archived complete data sets, 10 archived incomplete data sets, and, for the other four, it was not possible to tell whether the data set was complete or not. Only two people provided any meta-documentation at all and two of three people who provided analyzed data in databases did not provide any raw data. I give this example simply to say that abstract debates around the value and feasibility of secondary analysis need to engage with the reality of the datasets that are made available for this purpose. Although practices may improve over time, with support from organizations such as the UK Data Archive or US National Science Foundation, creating a complete dataset is more a labor of love than a realistic expectation. For development projects, it often relies on the un-costed labor of researchers in country or is done when they could be engaging in higher status activities such as writing. Quantitative researchers wanting to replicate the work of others face a similar challenge since it can take six to eight months of work to produce a dataset that is both replicable and similar to the dataset that was originally collected (Maren Duvendack, personal communication). Part of the problem is the increasing sophistication of econometric analyses, which critical quantitative researchers argue can conceal the quality of the underlying data (Angrist & Pischke, 2010). This is a concern that becomes particularly acute in the big-data era and in the next section I look at whether qualitative data analysis software strengthens or obscures claims to transparency and rigor.

3. Route to rigor? Qualitative data analysis software

One of the foci of this special issue is how new data collection (e.g. harvesting) and analysis techniques are being developed with the aid of new technologies able to process big data. Computer Aided Qualitative Data Analysis Software or CAQDAS has been around since at least the mid-1980s – the term was coined in 1991 – and has been made popular by programs such as NVivo⁶, Atlas t.i., and MAXQDA. The move towards larger data sets and the inclusion of functions such as auto-coding justify the time required to use it effectively. However, do CAQDAS improve the quality of data analysis? Their strengths relate to their speed (once the data has been coded), the ease of spotting

⁶ NVivo, or Nudist as it was then, was launched in 1999 and is now on its 11th version.

connections through the way the codes are structured, the potential to merge qualitative and quantitative data, and the avoidance of anecdotalism through the checking of the frequency of codes. They have some weaknesses, however, as they can distance analysts from the data, especially if they have not done the coding themselves. They make it easy to cherry pick quotes to support arguments, to take things out of context, and to quantify. Quantification may give the illusion of rigor, but it often takes us away from the purpose of applying qualitative research in the first place. As James (2012, p.7) observes, “the more interesting question may be not how many people said X with Y, but why they said it at all.” When focus group data are coded, they lose the dialogic element. And the narratives within interviews can also be fragmented through the coding process. This means that it is not always possible to see how data might be biased through the circumstances in which it was elicited (for example, the question immediately before the quotation). Although CAQDAS enable the extraction of coded quotations, most programs try to preserve context by allowing the analyst to return to the document the quotation came from and look at other documents within that case or set (e.g. a household or a social group). Programs such as Sensemaker and Python-based Text Data Mining tools extend CAQDAS into the realm of big data. They offer a different style of analysis as they look at patterns across fragments of text, abstracted from their original context, much as we would analyze quantitative data. This makes them well suited to analyzing tweets or other digital ‘micro narratives’ to identify cultural trends, but less able analyze context-rich cases where there are multiple types of information about an individual or in-depth interviews.

Some critics argue that analysis software removes the craft element of analysis (Mason, 2002; James, 2012). I advise my students to treat it as a data management tool and not let it do their thinking for them. Perhaps more importantly, it does not make analysis more transparent as much of the interpretation happens in the authors’ heads and is not captured in memos and notes. It does, however, act as useful shorthand in a journal paper for a systematic analysis has taken place, much in the way that claims of grounded theory used to do. Although textbooks provide some guidance on doing qualitative analysis, there is little reflection on the process outside of specialist journals. Below I briefly discuss some of the tensions within the analytical process, before looking at quality standards for qualitative research in the final section (section 4).

Analysis of any sort requires a reflexive, imaginative leap grounded in deep immersion in the data. Over time, James (2012) argues, we all become secondary analysts as the immediacy of our initial impressions recede. In both primary and secondary analysis, we need to ensure that the choices we make, while turning data into evidence through an inevitable process of selection and reduction, are not arbitrary, but arise from a close engagement with the data: “it is as profitable to ‘hang out’ in a

set of interview transcripts of conversations with people one does not know, as it is to hang out in their living rooms” (James, 2012, p.13).

Mason (2002) proposes that qualitative data analysis involves reading beyond the data. This enables us to locate narratives in a particular social context and draw on our empirical and theoretical knowledge as researchers to understand our data, rather than simply to report our findings. In relation to this process, the experience of not-being-there, whether as a result of team based research or secondary analysis, can even be profitable “through the continual dialogue that is involved in translating ‘the field’ for those who are not there by those who are there, even greater clarity and deeper insights can be engendered in team-based research” (James, 2012, p. 4)⁷. The emphasis placed by James on collaborative analysis is not always a part of team working, given that teams are often split into data collectors, analysts, and writers who are rarely located in the same institution, or country (Camfield, 2014). However, collaboration has the potential to produce more rigorous analysis as those involved need to defend their interpretations to others who know the data equally well (Thompson et al., 2012, Phoenix et al., 2016). It can be a challenge to create the conditions for this type of exchange given the hierarchies within research projects. Hossain and Scott-Villiers, this volume, observes that the US model of research frequently involves the use of graduate students and post-docs to conduct the fieldwork, which their supervisor writes up as paper or book. This division of labor and reward resonates with the experiences of researchers in the global South.

The potential benefits of collaborative working for rigor in qualitative analysis are illustrated by the work on data access and research transparency (DA-RT⁸) within North American Political Science. The DA-RT Guidelines on the provision of qualitative data alongside journal submissions make a helpful distinction between production and analytic transparency, which relates to Hammersley’s (2010) distinction between data as given and data as evidence. Production transparency addresses the process of data generation, and considerable guidance exists on how to make this transparent, as I discussed in the first half of this paper. Analytic transparency requires the author to provide evidence for their claims in a paper, for example, by linking to longer extracts of data that show how the ideas put forward developed. As I discussed in the preceding two sections, this requires not only reflexivity, but also a high level of documentation on analytical processes. Arguably, this is more likely to exist in team-based than solo analysis where what Geertz (1973) refers to as the “alchemy” of analysis may still hold sway. In the final section, I look at whether existing qualitative guidelines

⁷ Secondary analysts may also benefit from insights arising from the strangeness that anthropologists prize on first encountering a new field site.

⁸ See <https://www.dartstatement.org/> for guidelines to journal editors and other resources.

can help address some of the problems identified in the earlier part of the paper, specifically in relation to analysis.

4. Applying and assessing qualitative research guidelines

Debates over the desirability of qualitative data archiving and the danger that the models proposed for this might encode particular ways of understanding qualitative research are mirrored in debates over qualitative guidelines. In general, these are seen to be a good thing⁹, and particularly in the context of applied health research (Mays and Pope, 2000) and evaluation (Lub, 2015). However, there has been considerable debate over what they should contain (Hammersley, 2007) and their usability. For example, Barbour (2001) argues that they privilege minor technical details at the expense of the overall research design. Carter and Little (2007, p. 1316) propose transcending checklists with a broader definition of good quality qualitative research as “research that attends to [epistemology, methodology and method] and demonstrates internal consistency between them.” There is also a question about how guidelines might need to be adapted for secondary analysis. My experience reviewing the DFID-ESRC program, which I described earlier in the paper (see also Camfield et al., 2016), was that very little information was provided about who collected and analyzed data and this meant that the political economy dimension of ‘whose labor?’ was entirely obscured. This is not uncommon (see also Mauthner & Doucet, 2008). James (2012, p. 2) notes that,

in comparison to the amount of words that, over the years, have been dedicated to the ‘how to’ elements of the research process – how to design research, how to do this method or that, and how to organize one’s data ready for analysis – curiously, relatively little recent attention has been given to considering what is actually involved in the processes of interpretation.

The implications of this are that researchers cannot easily learn from each other’s work and the professional norms and standards for analysis are not communicated to the next generation (Corti & Fielding, 2015).

Within the field of international development, the most influential guidelines have been those authored by Spencer et al. (2003), which were commissioned by the UK government as a framework for assessing research evidence. Their report highlights the importance of conveying the “detail, depth and complexity (i.e. richness) of the data” (p.15). It does this by employing 18 questions relating to nine key areas: findings, design, sample, data collection, analysis, reporting, reflexivity

⁹ With some dissenting voices, e.g. Seale, 1999 who notes that interpretivist criteriologists have produced a set of bewildering criteria, Dixon-Woods et al, 2004, who suggests it stifles creativity, and Smith and Hodkinson, 2005 who flag the political dimension of setting criteria.

and neutrality, ethics and auditability. This is by far the most comprehensive of the available appraisal tools and has been used widely in applied research and evaluation, including by DFID (c.f. DFID How-to-note, 2014). It has, however, been critiqued by Torrance (2011, p. 574) for “presenting a counsel of perfection” that overlooks the “contingencies, political pressures and decisions that have to be made” as part of research. It is also hard to use effectively. In the online Appendix, I reproduce an analysis from Camfield et al. (2016) by applying the tool to the best-documented qualitative studies from the DFID-ESRC program. Despite the high level and quality of documentation, we did not have all the information that was required to complete the tool¹⁰. For example, section 5 on “How clear is the basis of evaluative appraisal?” is blank since there was no explicit reflection on how judgements were made, a problem highlighted in section 3 of this paper. Many of the requirements were only partially fulfilled; while data sources were triangulated internally, they did not bring in other “corroborating evidence to support or refine findings” or provide “clear discussion of how findings have contributed to knowledge and understanding.” There was also little “discussion of limitations on drawing wider inference (e.g. re-examination of sample and any missing constituencies).” The reluctance to discuss limitations, even where these might affect the conclusions drawn, is characteristic of all the studies looked at as part of the DFID-ESRC review and reflects a realistic assessment that future funding might be threatened by complete transparency. The combination of the level of detail required, the high expectations of research practice (informed by experiences in the global North), and the structural disincentives to transparency support Torrance’s conclusions on operability.

Below I present a summary of qualitative guidelines used in applied research (Table 1), drawing on Befani and Duvendack (2017) and supplemented with further searches of applied journals and websites. This shows firstly the dominance of health services research in their development (and in relation to this, a particular understanding of how research works) and, secondly, their increasing complexity over time. While Corti’s (2000) claim that there are no qualitative guidelines is no longer true¹¹, their ability to capture rigor within qualitative research is questionable. For example, almost all of the guidelines mention analysis, however, it is not given a great deal of weight relative to data generation, since it is usually only one of eight or more areas to consider. This is the case even in

¹⁰ I have had similar problems trying to use the excellent template for the INTERNATIONAL SOCIAL RESEARCH METHODS CASE STUDIES databank. This suggests that warranting quality by collecting ever more detailed information on process may not be realistic, or productive, if key details are no longer visible within the volume of information presented
http://www.restore.ac.uk/ISResMeth/Submission_of_CS/documents/ISRM%20Guidelines%20for%20Authors%20of%20Case%20Studies.pdf

¹¹ Their proliferation may be partly explained by the parallel growth in systematic reviews and meta-analyses (Duvendack et al., 2016) which require the material included to be assessed using common standards.

guidelines that are more focused on reviewing quality evidenced within research outputs than in providing principles for quality throughout the research process (e.g. Reynolds et al., 2011). It points towards a gap that I described earlier that while there are reflective articles on analysis in specialist journals (e.g. Thompson et al., 2012; Phoenix et al., 2016), if you are a critical development reader wanting to understand how a conclusion was reached, you will not be able to do this from the information provided. Referring again to the DFID-ESRC review, our analysis of outputs found that 82% of the qualitative outputs (45) do not describe their analytical approach, compared to 4% of quantitative outputs (2). Although this does not mean that the approach taken to the qualitative analysis was less systematic – reporting conventions are different in quantitative papers where the quality of the analysis is more important than the quality of the underlying data (Angrist & Pischke, 2010) – it makes it harder to see how data became evidence.

Table 1: Qualitative guidelines used in applied research, 1998 onwards

Authors	Focus	Application
Popay, J., Rogers, A. & Williams, G., 1998. Rationale and standards for the systematic review of qualitative literature in health services research.	Original article addresses appropriateness of methods and methodological soundness. Developed into a ten-point checklist by Attree and Milton (2006) addressing Background of research, Aims and objectives, Context, Appropriateness of design, Sampling, Data collection, Data analysis and findings, Reflexivity, How valuable or useful is the research, Ethics.	Research
Patton, M.Q. (2003) Qualitative Research and Evaluation Methods.	Provides a comprehensive set of guiding questions related to the appropriateness of the research design, fieldwork, ethics, analysis and reporting.	Evaluation
Spencer, L., Ritchie, J., Lewis, J. & Dillon, L., 2003. Quality in Qualitative Evaluation: A Framework for Assessing Research Evidence.	Provides 18 questions relating to 9 key areas: findings; design; sample; data collection; analysis; reporting; reflexivity and neutrality; ethics and auditability	Evaluation and commissioned government research
Dixon-Woods, M., Shaw, R.L., Agarwal, S., & Smith J.A. (2004). The problem of appraising qualitative research.	Provides a generic set of prompts relating to reporting, study design and execution.	Health services research
Long, A. F., & Godfrey, M. (2004). An evaluation tool to assess the quality of qualitative research studies.	Provides 34 questions across 4 key areas: Phenomenon studied and context; Ethics; Data collection, analysis and potential researcher bias; Policy and practice implications.	Health and social care
Walsh, D., & Downe, S. (2006). Appraising the quality of qualitative research.	Covers 8 key areas: Scope and purpose; Design; Sampling strategy; Analysis; Interpretation; Reflexivity; Ethical	Health services research

	dimensions; Relevance and transferability.	
Critical Appraisal Skills Programme (CASP). (2006)	Provides 10 questions relating to rigor, credibility and relevance.	Health services research
Tong, A., Sainsbury, P., & Craig, J. (2007) Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups,	Proposes the COREQ, a 32-item checklist for interviews and focus groups grouped under (i) Research team and reflexivity, (ii) Study design and (iii) Data analysis and reporting	Health services research
Daly, J., Willis, K., Small, R., Green, J., Welch, N., Kealy, M., & Hughes, E. (2007). A hierarchy of evidence for assessing qualitative health research.	Focuses on central methodological procedures such as defining a research framework, sampling and data collection, data analysis, and drawing research conclusions. Creates a hierarchy of qualitative research designs, which it proposes reflects the reliability of study conclusions.	'Evidence hierarchy' for assessing qualitative research within health.
Kitto, S.C., Chesters, J., & Grbich, C. (2008). Quality in qualitative research.	Looks at the extent to which qualitative reporting addresses Clarification and justification; Procedural rigor; Representativeness; Interpretative rigor; Reflexivity and evaluative rigor; and Transferability.	Health services research.
Tracy, S. J. (2010). Qualitative Quality: Eight "Big-Tent" Criteria for Excellent Qualitative Research.	Proposes eight markers of quality: Worthy topic, Rich rigor, Sincerity, Credibility, Resonance, Significant contribution, Ethics, and Meaningful coherence.	Research
Hannes K. (2011). Critical appraisal of qualitative research. In: Noyes J, Booth A, Hannes K, Harden A, Harris J, Lewin S, Lockwood C (editors), Supplementary Guidance for Inclusion of Qualitative Research in Cochrane Systematic Reviews of Interventions. Version 1. Cochrane Collaboration Qualitative Methods Group, http://cqrmg.cochrane.org/supplemental-handbook-guidance , downloaded 28/02/18.	Proposes assessing (i) adequacy of reporting detail on data sampling, collection and analysis, (ii) technical rigor of the study elements indicating methodological soundness, and (iii) paradigmatic sufficiency, referring to researchers' responsiveness to data and theoretical consistency.	Health services research
National Institute for Clinical Excellence (NICE) (2012). Appendix H Quality appraisal checklist – qualitative studies. Methods for the development of NICE public health guidance (third edition) (PMG4). https://www.nice.org.uk/process/pmg4/chapter/appendix-h-quality-appraisal-checklist-qualitative-studies , downloaded 24/02/18.	Proposes assessing theoretical approach (e.g. is qualitative approach clear and appropriate?), study design, data collection, trustworthiness (e.g. are the role of the researcher and the context fully explained?), analysis, ethics	Health services research

DFID (2014). How to note: Analysis of Qualitative Data in Evaluation and Research. Unpublished paper.	Assesses Transparency, Bias, Richness of data and presentation of diversity, Triangulation of analysis, Discussion of limitations, and Ethics.	International development
---	--	---------------------------

Source: Authors' own

5. Conclusion

In the paper, I have argued that rigor is an ethical issue, especially in a field such as development that is both applied and resource constrained. Researchers have a responsibility to ensure that their work is rigorous because failure to do so has real consequences. I note that while there have been improvements in documenting the production of research and sharing data, less attention has been paid to analytical transparency. This includes how analysis and interpretation were done (and by whom), and also the evidential basis for claims and how these claims are “warranted” (Manski, 2011:25). In large international projects, data sharing represents the status quo, although this is rarely acknowledged. This means that data users are often remote from those who generated the data. Without full information on how this was done and the constraints this places on any analysis (e.g. due to the particularity of the sample), they may struggle to use it responsibly. Even though detailed descriptions of analyses are usually confined to specialist journals, understanding how the analysis was done – how data was turned into evidence - is important in judging its quality and the weight that can be placed on its conclusions. This is especially the case as new actors such as program evaluators and quantitative researchers use and assess qualitative methods.

In looking at potential solutions, I highlighted the proliferation of guidelines to ensure the quality of qualitative research, especially within health services research. These were developed in part to address these threats to rigor. However, I also suggested that, in relation to analysis, these provide little practical support, especially in overcoming the common and hierarchized separation between the generation and interpretation of data. This can lead to the loss of context, which is arguably one of the main reasons why one collects qualitative data rather than quantitative. I also note that context can also be lost through CAQDAS, which supports practices previously critiqued in quantitative research such as data mining and cherry picking of quotes to support an argument. This is a particular problem with newer variants of CAQDAS that are being developed to work with qualitative big-data and then applied to qualitative data analysis more broadly, although I don't discuss this fully here.

In relation to preserving context, I looked at the guidance given to data archivers and the level of documentation needed to be able to engage with the data as though you had collected it yourself. While certainly comprehensive, I suggested that outside archival demonstration projects such as

Timescapes¹², the guidance might not be realistic. Within development, the expectations it creates puts considerable burdens on staff in-country. This has ethical implications since a visible aspect of development research is how few Southern researchers are lead authors on publications, especially where these are qualitative (Camfield, 2018).

Ironically, guidance for secondary analysis largely fails to address the question of analysis, which I argue presents the greatest threat to rigor in development research. The reason for this is that similar to the qualitative guidelines discussed in the final section its primary focus is the generation of data, rather than evidence. I propose that moving towards a more participatory form of analysis, where analysts work in teams drawn from across the project, might produce a richer and more sensitive analysis. It would make it possible to combine vital contextual knowledge, including how the data was generated, with an awareness of the conventions of academic writing and broader debates within related literature or international policy circles. More importantly, and with attention to hierarchies and power imbalances that occur across research projects, it would create an opportunity to engage critically with the analyses produced and consider other possible interpretations. This process of sense-checking across the team would improve the quality of the analysis in the same way as feeding back to the communities where the data was generated. Not only is feeding back an important ethical commitment, but errors are instantly picked up and analysts realize that their authoritative interpretation is inevitably partial. The impact that a partial analysis might have on the people who provided the data underlines the importance of considering rigor in qualitative analysis as an ethical issue.

¹² <http://www.timescapes.leeds.ac.uk/>, downloaded 06/03/18.

References

- Angrist, J.D., & Pischke, J-S. (2010). The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics. *Journal of Economic Perspectives*, 24(2), 3-30.
- Attree, P., & Milton, B., (2006). Critically appraising qualitative research for systematic reviews: defusing the methodological cluster bombs. *Evidence & Policy*, 2 (1), 109-126.
- Bakewell, O. (2007), Breaching the Borders between Research and Practice: Development NGOs and Qualitative Data, In Matt Smith (ed.) *Negotiating Boundaries and Borders* (Studies in Qualitative Methodology, Volume 8, pp. 217-238). Emerald Group Publishing Limited.
- Barbour, R. (2001). Checklists for improving rigour in qualitative research: a case of the tail wagging the dog? *BMJ*, 322(7294), 1115–1117.
- Befani, B., & Duvendack, M. (2017). High Quality Evaluations: Criteria, Standards and Proposal for a Scoring System. BEIS, unpublished paper.
- Camfield, L. (Ed.). (2014). *Methodological Challenges and New Approaches to Research in International Development: a critical review*. Palgrave Macmillan, UK.
- Camfield, L. (2018). Methodologies in development studies. In: Baud, I, Basile, E and Kontinen, T. (eds). *International Development Studies*. London: Palgrave Macmillan.
- Camfield, L., Duvendack, M., & Monteith, W. (2016). ESRC-DFID Joint Fund for Poverty Alleviation Research Evidence 'Synthesis Research Award (ESRA) for Research Methods', Report, International Development: University of East Anglia, downloaded 27/11/17 Retrieved from <http://www.esrc.ac.uk/files/research/international/research-methods-esra-report/>
- Camfield, L., & Palmer-Jones, R. (2015). Ethics of analysis and publication in international development. In: Nakray, K, Alston, M, Whittenbury, K. *Social Science Research Ethics for a Globalizing World: Interdisciplinary and Cross-Cultural Perspectives* (chapter 15). London: Routledge.
- Camfield, L., Duvendack, M., & Palmer-Jones, R.P. (2014). Things you Wanted to Know about Bias in Evaluations but Never Dared to Think. *IDS Bulletin*, 45(6), 49-64.
- Carter S.M., & Little M. (2007). Justifying knowledge, justifying method, taking action: epistemologies, methodologies, and methods in qualitative research. *Qualitative Health Research*, 17(10), 1316-28.

Clifford, J., & Marcus, G. (Eds.). (1986). *Writing Culture*. Santa Fe, NM: School of American Research.

Copestake, J., Allan, C., van Bekkum, W., Belay, M., Goshu, T., Mvula, P., Remnant, F., Thomas, E. & Zerahun, Z., (2016). Managing relationships in qualitative impact evaluation to improve development outcomes: QuIP choreography as a case study. Working Paper. Bath, U. K.: Centre for Development Studies, University of Bath. (Bath Papers in International Development and Wellbeing; 2016/49)

Corti, L. (2000). Progress and problems of preserving and providing access to qualitative data for social research - the international picture of an emerging culture. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 1(3).

Corti, L., Fielding, N.F. & Bishop, L. (2016) Editorial for Special Edition, Digital Representations: Re-Using and Publishing Digital Qualitative Data, SAGE Open.

<http://dx.doi.org/10.1177/2158244016678911>

Duvendack, M., Palmer-Jones, R., & Reed, W. R. (2017) What Is Meant by “Replication” and Why Does It Encounter Resistance in Economics?. *American Economic Review*, 107 (5), 46-51.

Eriksen, T. H., & Schober, E. (2017) *Identity Destabilised: Living in an Overheated World*. Department of Social Anthropology, University of Oslo.

Fielding, N. (2004) Getting the most from archived qualitative data: epistemological, practical and professional obstacles. *International Journal of Social Research Methodology*, 7(1), 97-104.

Freeman, D. (1983). Margaret Mead and Samoa: The Making and Unmaking of an Anthropological Myth. Harvard University Press, M.A.

Geertz C (1973). *The Interpretation of Cultures*. London: Hutchinson.

Greckhamer, T., & Cilesiz, S. (2014). Rigor, Transparency, Evidence, and Representation in Discourse Analysis: Challenges and Recommendations. *International Journal of Qualitative Methods*, 13(1):422-443.

Guba, E.G., & Lincoln, Y.S., 1985. *Naturalistic inquiry*. Beverly Hills, CA: Sage Publications.

Guijt, I., & Roche, C. (2014). Does Impact Evaluation in Development Matter? Well, It Depends What It's For! *European Journal of Development Research*, 26, 46-54.

Hammersley, M. (2010). Can we re-use qualitative data via secondary analysis? Notes on some terminological and substantive issues. *Sociological Research Online*, 15(1).

- Hammersley, M. (2007). The issue of quality in qualitative research. *International Journal of Research & Method in Education*, 30, 287–305.
- Irwin, S., & Winterton, M. (2012). 'Qualitative Secondary Analysis and Social Explanation'. *Sociological Research Online*, 17(2),4.
- Irwin, S., & Winterton, M. (2011). Timescapes data and secondary analysis: working across the projects. Timescapes Working Paper 5.
- James, A. (2012). Seeking the analytic imagination: reflections on the process of interpreting qualitative data. *Qualitative Research*, 13(5), 562–577.
- Johnson, S., & Rasulova, S. (2017). Qualitative research and the evaluation of development impact: incorporating authenticity into the assessment of rigour. *Journal of Development Effectiveness*, 9(2), 263-76.
- Lub, V. (2015). Validity in Qualitative Evaluation: Linking Purposes, Paradigms, and Perspectives. *International Journal of Qualitative Methods*, 14(5), 1-8.
- Manski, C. F. (2011). Policy Analysis with Incredible Certitude. *The Economic Journal*, 121, F261-289.
- Mason, J. (2002). *Qualitative researching* (2nd edition). London: Sage.
- Mauthner, N.S., & Parry, O. (2009) Qualitative data preservation and sharing in the social sciences: On whose philosophical terms? *Australian Journal of Social Issues*, 44(3), 291-307.
- Mauthner, N. S., & Doucet, A. (2008). 'Knowledge once divided can be hard to put together again': an epistemological critique of collaborative and team-based research practices. *Sociology*, 42(5), 971-985.
- Mays, N., & Pope, C. (2000). Assessing quality in qualitative research. *British Medical Journal*, 320, 50–52.
- Ottenberg, S. (1990). Thirty years of fieldnotes: Changing relationships to the text. In R. Sanjek (Ed.), *Fieldnotes* (pp. 139–160). Ithaca, NY: Cornell University Press.
- Palermo, T. (2017) Are randomized controlled trials bad for children?
<https://blogs.unicef.org/evidence-for-action/are-random-control-trials-bad-for-children/>,
downloaded 26/02/18.

- Phoenix, A., Brannen, J., Elliott, H., Smithson, J., Morris, P., Smart, C., Barlow, A., & Bauer, E., (2016). Group Analysis in Practice: Narrative Approaches. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 17(2), Art. 9.
- Reynolds, J., Kizito, J., Ezumah, N., Mangesho, P., Allen, E., & Chandler, C. (2011). Quality assurance of qualitative research: a review of the discourse. *Health Research Policy and Systems*, 9(43):1-10.
- Savage, M. (2010) *Identities and social change in Britain since 1940: the politics of method*. Oxford, UK: Oxford University Press.
- Seale, C. (1999). Quality in qualitative research. *Qualitative Inquiry*. 5(4), 465-478.
- Smith, J.K., & Hodkinson, P. (2005). Relativism, criteria, and politics. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 915-932). Thousand Oaks, CA: Sage.
- Temple, B. (2005). Nice and Tidy: Translation and Representation. *Sociological Research Online*, 10(2) doi:10.5153/sro.1058.
- Thomson, R., Moe, A., Thorne, B., & Nielsen, H.B. (2012). Situated Affect in Traveling Data: Tracing Processes of Meaning Making in Qualitative Research. *Qualitative Inquiry*, 18(4), 310-322.
- Torrance, H. (2011). Qualitative research, science and government. Evidence, criteria, policy and politics. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (4th ed., pp. 569-80). Thousand Oaks, CA: Sage.
- Wilson, S. (2014). Using Secondary Analysis to Maintain a Critically Reflexive Approach to Qualitative Research. *Sociological Research Online*, 19 (3), 21 DOI: 10.5153/sro.3370