



HHS Public Access

Author manuscript

Int J Qual Methods. Author manuscript; available in PMC 2024 April 25.

Published in final edited form as:

Int J Qual Methods. 2024 ; 23: . doi:10.1177/16094069241236268.

Remote and Equitable Inductive Analysis for Global Health Teams: Using Digital Tools to Foster Equity and Collaboration in Qualitative Global Health Research via the R-EIGHT Method

Jason Johnson-Peretz¹, Titus O. Arunga², Joi Lee¹, Cecilia Akatukwasa³, Fredrick Atwine³, Angeline Onyango², Lawrence Owino², Carol S. Camlin^{1,4}

¹Department of Obstetrics, Gynecology & Reproductives Sciences, University of California, San Francisco (UCSF), San Francisco, CA, USA

²Kenya Medical Research Institute (KEMRI), Kisumu, Kenya

³Infectious Diseases Research Collaboration, Kampala, Uganda

⁴Department of Medicine, University of California, San Francisco (UCSF), San Francisco, CA, USA

Abstract

Qualitative methods encompass a variety of research and analysis techniques which have the common aim of uncovering what cannot be captured numerically through the quantification of data. For qualitative analytical methods in the interpretivist tradition (e.g. grounded theory, phenomenological, thematic, etc), inductive coding has become a mainstay but has not always lent itself to collaborative, remote team-based data interpretation among qualitative and mixed-methods clinical researchers. Finding ways to speed the inductive coding process without

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

Corresponding Author: Jason Johnson-Peretz, Department of Obstetrics, Gynecology & Reproductives Sciences, University of California, San Francisco, Mission Hall: Global Health & Clinical Sciences Building, 550 16th Street, 3rd Floor, San Francisco, CA, USA. Jason.johnson2@ucsf.edu.

Definitions

Approach(es)

Holistic consideration of the data sets, research questions, and desired products involved in a research project, that when taken together inform the choice of analytical method to examine said data. Includes decisions about how to meet, delegate tasks, and come to a consensus about next steps or open questions.

Code/Coding

Categorizing segments of data with a short name that simultaneously summarizes and accounts for each piece of data.

Codebook

A list of codes, limited in number, which researchers apply to all transcripts or written material in a dataset. The codebook keeps researchers on the same page with respect to the range of ideas to be marked and the terms used to do the marking of specific segments of data.

Inductive coding

An interpretivist qualitative approach describing the process of defining what segments of qualitative data are about, in order to draw out emergent findings.

Method

The specific technical process chosen to analyse the data in question.

Declaration of Conflicting Interests

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

sacrificing rigour while remaining accessible to geographically dispersed teams remains a priority. This is especially crucial in global health partnerships where on-the-ground researchers may have less input into codebook development compared to in-the-office researchers. We describe a newly-developed, digital approach that integrates findings from our qualitative team, which we call R-EIGHT (Remote and Equitable Inductive Analysis for Global Health Teams). *The technique we developed a) speeds the process of inductive coding as a team, b) visually displays interpretive consensus, and c) when appropriate fosters streamlined integration of inductive findings into codebooks.* Because it involves all team members, our approach helps break the divide between in-office and on-the-ground teams, fostering integrated and representative contributions from all globally-dispersed team members.

Keywords

qualitative methods; remote teams; global health; inductive coding; equity

Introduction

Qualitative methods encompass a variety of research and analysis techniques which have the common aim of uncovering what cannot be captured numerically through the quantification of data; they instead follow an ‘interpretive paradigm’ (Castleberry & Nolen, 2018; Denzin, 2017; Knoblauch, 2005). Examples of appropriate qualitative targets include research seeking to uncover the mechanisms of action by which an intervention works, the real-world experience(s) of a particular intervention, and the lived contexts which affect participants in the intervention.

Among the more popular qualitative methods today, inductive analysis and grounded theory have become mainstays in qualitative and mixed-methods clinical research (Charmaz, 2014; Vaismoradi et al., 2013). Inductive analysis, of course, is not limited to grounded theory alone, and codes applied during inductive analysis can be used in several analytical traditions, whether to develop codebooks, thematic tables, or theory.

These qualitative methods, however, can be time-consuming and too open-ended from the perspective of quantitatively trained teammates, and they have not always easily lent themselves to remote teamwork. Further, inclusion of local expertise, ideally always the practice in geographically dispersed, university-based clinical global health research, is necessary for robust data interpretation. Simultaneously making the process accessible to geographically-dispersed teams while finding ways to speed the process of thinking and interpretation without sacrificing rigour remains a priority. Innovation, in the sense of adapting existing methods through creative use of tools already present in digital software or platforms, is necessary – especially to facilitate adoption across remote-work teams (Caliandro & Gandini, 2016; Thunberg & Arnell, 2022; Wiles et al., 2011).

Turning specifically to codebook development for qualitative researchers, multiple perspectives and methods are inherent to the analytic process and reflect the working dynamic that fuels global health collaborations. Without integrating these various perceptions into a functioning product, the risk of omitting valued contributions from all

team members threatens the twin goals of achieving equity in global health research and of designing research for a balanced representation of the participants studied (Gallagher & Kim, 2008; Holeman & Kane, 2020; Pratt & Hyder, 2018). This is especially clear when in-office researchers from outside the participants' own country make the decision to omit the insights and contributions from teammates on the ground, or fail to solicit such insights altogether. Adapting existing methods thus has a moral valence as well as a technological one (Denzin, 2017; Pratt & Hyder, 2018; Wiles et al., 2011).

Prior Approaches

After gathering data via in-depth interviews, focus group discussions, and observation, qualitative analysis sometimes proceeds by applying codes from a previously developed codebook (an *a priori* codebook) to the interview or focus-group transcript, often using coding software like Dedoose or Atlas.ti to facilitate the coding of large data sets. Codes are short names given to segments of data that simultaneously summarize and account for each piece of that data. *A priori* codebooks typically focus on topics related to the specific research questions that informed the creation of the initial interview guide or study proposal and may reflect theoretical constructs that underlie the premises of the research. In a purely deductive approach, the codebook would not be further refined on the basis of inductive coding and may not reflect the decisions or insights from teammates on the ground, that is, the researchers who actually live and work among the study population.

Inductive analysis, however, proceeds from the 'ground up', drawing out meaning through line-by-line interpretation of what the participants say, and then abstracting those findings into increasingly refined codes from which data interpretation follows (Charmaz, 2014; Lipscomb, 2012; Timmermans & Tavory, 2012; Vila-Henninger et al., 2022). This is useful particularly when doing non-hypothesis-based research in order to identify novel findings. It also helps draw out themes important for the community members being interviewed which may be at variance with the assumptions of the team which proposed the initial research questions. In semi-structured interviews which can elicit unexpected narratives and perspectives from informants/participants, inductive coding allows for researchers to account for those novel findings and incorporate them into a codebook for further identification of similar ideas offered by other participants. Inductive analysis has been appropriate for our team also because we have smaller data sets of participants ($10 < N < 150$). When we have a dataset at the larger end of that range, we perform inductive coding on a sub-set of interview transcripts and then integrate those findings with an *a priori* codebook for application to the remaining transcripts.

Charmaz describes two phases of the inductive coding process used in grounded theory: an initial phase with line-by-line coding of the data to highlight actions and processes, and a focused, selective phase for refinement of those codes; she adopts the term axial coding to refer to the categorization of codes and subcodes (Charmaz, 2014). Similarly, Saldaña describes this process as first-cycle and second-cycle coding, wherein "the primary goal during second cycle coding is to develop a sense of categorical, thematic, conceptual, and/or theoretical organization from your array of first-cycle codes" (Saldaña, 2016).

This inductive approach is particularly useful in analysing unstructured and semi-structured interviews, where unexpected findings emerge that were not conceived of or accounted for in *a priori* codebooks. These inductively derived codes can be applied to transcripts through the creation of an inductively-derived codebook (uploaded into coding software), or integrated into an *a priori* codebook for a more inclusive and layered analysis of the data, in what Charmaz calls axial and what Saladaña calls second-cycle coding. In this paper, we use the second option, that of integrating *a priori* and inductively-derived codebooks for remote and geographically dispersed teams, involving both in-office and on-the-ground team members.

Revisiting Approaches to Coding in Global Research Contexts

When it comes to ground-up analysis, inclusive participation in the development of the codebook through incorporating inductively-derived codes increases the validity and rigour of findings by ensuring interpretation is checked by those with local cultural understandings (Karnilowicz et al., 2014; Salmen et al., 2022). (A distinction between rigour and validity is implied, though not explicitly stated, by Salmen's and Karnilowicz's articles). In other words, including those with local expertise on the interpretive team strengthens the validity of study results through *cultural rigour* (Lock et al., 2021) which "affirms the value of experiential knowledge and stresses a collaborative process" (Leung et al., 2004). This helps support the *relevance* of the codes being investigated as well as the potential *reach* into the community of those products ultimately derived from analysis of excerpts so coded (Balazs & Morello-Frosch, 2013). Fundamentally, such participation keeps the orientation of the research interpretation process centred around those who know, are a part of, or who otherwise participate in the multi-cultural context of the communities involved (Minkler, 2004). To quote Israel et al., (1998), "community is characterized by a sense of identification and emotional connection to other members, common symbol systems, shared values and norms, mutual – although not necessarily equal – influence, common interest, and commitment to meeting shared needs." Any of these elements can be overlooked or completely missed when codes – the fundamental *terms of analysis* – are created exclusively by researchers outside of or only peripherally engaged in the research participant communities (Israel et al., 1998).

Obviously, this process can be time-consuming; finding ways to move it forward efficiently without sacrificing accuracy is helpful for teams, especially in the context of multidisciplinary research projects including large community trials with both quantitative and qualitative study aims. Secondly, being able to address the objections of reviewers who seek to quantify qualitative methods is often important for qualitative researchers on clinical teams, as we repeatedly encounter reviewers who do not understand the merits of systematic, replicable, qualitative research as such. Often kappa coefficients or other quantitative methods to determine inter-coder reliability become frustrating endeavours for qualitative researchers when some team members are 'minimal coders', (or, colloquially, 'splitters') who highlight only brief excerpts for coding, while others are 'maximal coders' (colloquially, 'lumpers') who highlight interviewer questions or the larger context of an excerpt or quote, plus the specific line warranting a code (Deterding & Waters, 2021; Hemmler et al., 2022). A mix of minimal and maximal coders on a team will throw off

kappa coefficients, while any team who reviews such discrepancies often finds that everyone had already reached consensus about what code to apply and how to interpret the codes; they simply do not code the same amount of text for quantifiable measurement to accurately reflect consonance (and replicability) in a quantitative context. Finding alternative ways to address that objection and highlight consensus as the metric, rather than quantity, is therefore also useful for qualitative researchers.

A Global Test-Case for a New Method

The qualitative arm of the SEARCH-SAPPHIRE study has provided a useful case study in overcoming the above challenges. The Sustainable East Africa Research in Community Health collaboration is an academic research consortium based in Kenya (Kenya Medical Research Institute – KEMRI), Uganda (Makerere University/Infectious Disease Research Collaboration – IDRC), and the United States (University of California, San Francisco – UCSF), conducting a series of community-based randomized controlled trials sharing the ultimate aim of ending the HIV epidemic and improving the health of rural eastern African communities (Havir et al., 2019). The collaboration includes quantitative- and qualitative-trained researchers across multiple disciplines in all locations, who by necessity meet remotely, and only occasionally meet in-person. In our case, the ‘on-the-ground’ qualitative team conducts semi-structured interviews or participant observation exercises in the field, then translates, transcribes, and uploads those interviews or field notes in the office. They have pre-existing software training (e.g. MS Office, Dedoose), and time is built into their schedules for coding transcripts individually. In the course of our collaboration, we have developed tools to facilitate communication and team-based analysis of our qualitative data.

Our previous efforts to streamline qualitative analysis relied on a rigorous yet imbalanced approach to reviewing transcripts, and developing and revising coding frameworks: initially, the SEARCH qualitative teams developed coding frameworks that were simple in structure, relying primarily on broad (or “parent”) codes that could easily be applied consistently by team members with varying levels of training and expertise across regions. Consequently, the interpretation of coded data, the ‘indexing’ step, involved a heavier task of analyzing less-refined sets of data. To foster inclusivity at this stage required in-person meetings for collaborative review and discussion of extracted sets of coded data. The structure of these meetings was constrained by the budgetary and scheduling limits inherent to funded research: one to two U.S.-based investigators traveled to eastern Africa to meet with multi-person teams, whose time for analysis workshops was further constrained by data collection schedules.

Such structural constraints highlighted the power imbalance that has been historically pervasive in global health collaborations, and limited the time and attention that field-based team members— with their valuable local cultural expertise and proximal access to study participants— were able to contribute to the consolidation of the key emergent themes and findings. Over time and with additional training and planning, the team developed more refined coding frameworks with iterative stages of data review and development of hierarchical inductive codes, reducing the burden of a second stage of interpretation that

required in-person meetings for inclusivity. Further, to address the structural barriers to full and equitable participation in the interpretive process, we worked on an integrated approach involving iterative reviews of in-person meeting notes via Zoom and offline Word document tracking. Our analytic approach was intended to be broad enough to allow the application of either grounded theory or framework analysis methods. Whichever method chosen, our approach is useful for inductive coding work as a dispersed team.

The purpose of this paper therefore is to describe a method we used to transfer former on-the-ground workshops to digital platforms. We call this the R-EIGHT (Remote and Equitable Inductive Analysis for Global Health Teams) method. In the process, we discovered the digital format not only streamlined coding and post-coding discussions but also provided us with a clearly documented ‘paper trail’, while facilitating the integration of inductive codes into *a priori* codebooks. We want to share this method in the hope that it will help other geographically-dispersed teams continue to maintain high and replicable standards of qualitative analysis while addressing imbalances between in-office and on-the-ground team members in global health partnerships.

Methods

The technique we developed is fairly straightforward. It consists of eight steps. First, (1) participant interviews are transcribed, during which time (2) a team member develops an *a priori* codebook from existing interview guides and study protocols. Then, (3) each team member is given a transcript or set of transcripts for initial or line-by-line coding using the ‘comment’ function in MS Word. (4) The team members then send their coded or commented-upon transcripts to a central teammate responsible for merging the commented-upon documents. This will produce a document in which team-mate comments for a given section all appear together in a side column. Next, (5) the team meets to review the comments, which are conveniently grouped together around the relevant text, to foster discussion and compare the language each coder is using. From there, (6) the team compares these inductively-derived codes with any *a priori* codes (if used) in axial or second-cycle coding to develop focused codes, in order (7) to define the range of application of an *a priori* code, and to add new parent and child codes to the codebook as necessary. The team can also discuss any misperceptions in coding or interpretation of a relevant excerpt as well, correcting the biases or simple lack of experience an in-office coder might have compared to an in-the-field coder. Step (8) is the finalisation of the codebook for coding the remaining data sets. Below, we describe each step of the process in greater detail.

Step One: Assembling Starting Materials

To use this process, the transcribed interviews to code must be in Word format. While other platforms like Google docs have commenting capability, the purpose of using Word is to maintain blinding among all coders so that one person’s codes – including the language of those codes – do not influence those of another team member when doing inductive, line-by-line coding. (This is also why we do not use the comment function in a shared Google document: those comments are visible to everyone, and we wanted an initial blinding of all reviewers in order to ascertain our initial degree of consensus.)

Our interviews are conducted in one of four languages: English, Swahili, Dholuo, and Runyankole. These audio-recorded interviews are all translated into English (an official language shared by both Kenya and Uganda) during the transcription process and then reviewed by native speakers of the relevant languages. We then proceed to analysis with the English-language versions of the interviews.

Step Two: Theoretical Codebook

Before beginning coding, one team member will have prepared a codebook of *a priori* codes using MS Excel. We derive the *a priori* codebook directly from semi-structured interview guides. Questions in our semi-structured interview guides typically encapsulate a particular theme which the team wants to capture in a code, and the questions therefore easily furnish *a priori* codes based on the expected participant responses to those particular questions.

The *a priori* codebook template has four columns, one each for the broad (parent) codes, child codes, notes, and examples. (See Figure 1) We also sometimes include a column for the particular interview guide prompt or the specific participant group to which the code most often applies.

The *a priori* codebook is not shared beforehand, to limit priming team members to those codes. (All coders, however, would likely be familiar with the interview guide.)

Step Three: Select and Distribute Sample Transcripts to Team Members for Initial Coding in MS Word

Once the transcripts and codebook are prepared, a central team member emails selected transcripts which will be used to derive inductive codes to all team members for individual line-by-line coding. Ideally, several transcripts will be used for pilot coding and discussion to create a more inclusive codebook which reaches code saturation. We assigned two transcripts (transcript A and transcript B) to two groups of four team members each (team A and team B). We recorded who was assigned which transcript, as this becomes important in Step 4.

Once all members have received the initial set of transcripts, each member individually codes the transcript line-by-line using the 'comment' function in Word. (The comment function can be found in the ribbon at the top of the page, under the 'Review' tab.) The coder selects a portion of text, goes to the comment function and clicks 'new comment'. The coder can then input a description of the highlighted text. The description format we use follows the first stage of Charmaz's two-stage approach to inductive coding, which she refers to as 'initial coding', slightly nuanced by the particular social science disciplines of each team member. (See Figure 2)

All team members code in the same language (English), since the transcripts are translated into that shared language, and because the resulting codebook is for the entire team's use. While English is a second language for some team members, we all attended English-language schools.

It is important to ensure the version of Word the coder uses is set up to track each user's name, else a generic 'Microsoft User' will appear in the comment header when these documents are subsequently merged. (To change this setting on a Mac using Word version 16, go to the File > Properties menu. A smaller pop-up window will appear. From there, select the 'Summary' tab. In the 'author' box, input the coder's name.)

During the initial individual coding task, team members closely study fragments of data – words, lines, segments, and incidents – for their analytic import, often applying gerund phrases to each line to code the action or process described. From time to time, we may adopt our participants' telling terms as codes, (or "*in vivo*" codes) in order to remain as open and close to the data source as possible. Codes tend to be short and stick closely to the data, showing actions and tracing the progression of events from the participant's point of view. For example, the codes cover participants' feelings, what happened, and how the participants explain their interpretation of events.

Initial open coding continues the interaction that interviewers shared with participants while collecting data but brings additional team members into that interactive analytic space. The codes at this stage portray meaning and actions to highlight new questions the team might have on the subject based on unexpected participant reports, and serve to provide a set of ideas that team members can later refer to, in Charmaz's second stage of 'focused' coding.

Team members then send the coded transcript back to the central team member who is responsible for the next step, merging the comments into a single document.

Step Four: Merge the Coded Documents

As the central team member receives the coded transcripts from individual coders, that teammate can begin to merge them. It is important to merge only those documents which have the same base text. If more than one transcript was used in coding, *do not merge transcripts from different interviewees!*

We merge the documents only two at a time. To merge them, select a base document, for example from 'Coder 1' (C1). Go to the ribbon at the top of Word, select the 'Review' tab, and then select the 'Compare' function. From there, the user has a choice to either 'compare' or 'combine' documents. While either could theoretically be used, we opted to use the 'combine' function.

When the 'combine' function is selected, Word prompts the user to choose a primary document to start the merging process, for example, Coder 1's commented-upon transcript, and a revised document, which in this instance would be the transcript commented on by Coder 2 (C2). Press OK. Word will then create a new, untitled document. Using the 'save as' function, name this document using the initials of the team members whose comments have now been combined into one document. The format we use is *transcriptA_C1_C2*, where C1 and C2 stand for 'Coder 1' and 'Coder 2' respectively. When Coder 3's transcript arrives (C3), the 'primary document' to use when combining documents will be the *transcriptA_C1_C2*, the revised document will be C3, and the resulting new document will

be saved as *transcriptA_C1_C2_C3*. It is helpful to save all these documents in one folder for ease of access.

The combined comments can often be in a font point too small for some people to read, especially if this document is shared via video conferencing platforms. Formatting may therefore help increase readability for the team. To increase the font size for the comments column in the final document, click on a comment (usually appearing on the right-hand side of the page) to open a scrollable column listing all revisions to the document on the left-hand side of the page. The merged comments will be grouped together in this 'revisions' column.

Once the revisions column is open and some text there is selected, click on the 'Home' tab in the ribbon at the top of the page. Go to the 'Styles' function and select 'styles pane'. This will open the styles menu to the right of the document. The box 'current style' should already read 'Comment Text'. From here, the user can select 'New Style'. This opens a new window, with an option to change the point size of the font. Select a point size readable by all team members. Close the window. The Revisions column on the right side of the pane should now have a larger typeface. Save the document.

Once complete, the central team member can send a copy of the final combined document to all coding team members for reference.

Steps Five through Six: Discuss the Merged Comments to Produce Focused Codes

Once all coder comments have been merged, we set a time to review these open codes as a team using Zoom or other video-conferencing platforms which have a screen-share capability. With the document's 'Revisions' column open and shared on screen, we proceed down the comments, section by section, and discuss the findings. We do this in tandem with integrating inductive codes into an *a priori* codebook (step 6).

If more than one team has coded transcripts, several options are available for these discussions. Teams could break out and discuss their own transcript together, going line by line and looking at how each coder commented upon the section. Alternatively, the entire team together could view a particular transcript and its comments, regardless of whether this was the transcript some members were assigned. The benefit of the latter approach is that members who had not seen that transcript can add insights from their experience of another interviewee; but this is also a drawback, in that it may slow down discussion when the second transcript is not visible on screen.

When reviewing the grouped comments, pay attention to similarities and divergences in the language each team member used to code a section of text. Differences in the language can foster discussion about what words should be adopted for the final or focused code name that will trigger the use of that particular code by all team members. We found it fascinating when multiple team members used the exact same language to describe a section of text; this fostered confidence in our ability to code passages similarly – and seemed a much more relevant way of gauging inter-rater reliability than a quantitative analysis based on the amount of text captured in any given excerpt.

Beyond language, we attend to consonance in ideas, tracking how much each team member is ‘on the same page’ even when the exact language differs (Figure 3). This can help identify teammates who are especially strong at succinct language; they may be particularly adept at choosing code names for the final codebook. Consonance of ideas can also help pinpoint areas in transcripts which might be prone to cause confusion in the future, in terms of theme, meaning, and interpretation. This occurs especially when finer distinctions need to be made or two code names overlap in terms of concepts. Similarly, when ideas about a particular section of text diverge, this can either expand team insights around data interpretation and further theoretical development. Divergences can also reveal which team members’ coding skills may need extra coaching or practice.

Sometimes only one team member will have commented on a sentence. In this case, like with divergent comments, the team can discuss whether the comment is a ‘valid’, ‘accurate’, or ‘insightful’ interpretation and whether that sentence ought to receive a code. Such discussions are especially important when some teammates are more familiar with the nuances of cultural expressions and the language of the participant, or when a teammate can recall the actual context and rapport of the interview under discussion (e.g. the teammate was the interviewer).

It can also be intellectually exciting when team members come from different theoretical or social science backgrounds and code the same line differently based on their discipline’s particular theoretical orientations. Often these conversations can spark further insights about themes to attend to when coding, additional cross-disciplinary papers to produce, and the ways in which the same words have different resonances within disciplines. In general, this sort of discussion helps teach team members about their particular disciplinary out-looks, and allay the potential for cross-communication or use of same-terms-different-meanings that can happen between social science disciplines.

We advise coding early in the research process to see where it takes the team as we proceed. This is a perfect time to have a team member start a coding memo document (distinct from post-interview field note memos) for the memo-ing phase of Charmaz’s method, in order to provide a reference document for any potential questions a future transcript might pose around what codes the team should apply.

Step Six: Develop Focused Codes

Building on step five and similar to step eight when the codebook is finalised, the team members discuss each set of comments from the merged document, and check whether those comments fit under an already existing code in the codebook. During this second-cycle, focused coding, we choose codes with the most analytic breadth and ability to subsume several initial codes. Sometimes this code comes from the *a priori* codebook, while at other times it emerges from the initial inductively-derived codes. This is where theoretical integration begins, as the focused codes begin to provide an analytical frame for future products.

As a team, we decide what the code will include and exclude: how far can any particular code reach? Initial codes are compared for their similarity, their impact, or their centrality;

sometimes similar but distinct codes are kept as child codes of a parent code whose thematic breadth includes both child codes, particularly if we are uncertain at the outset whether a child code warrants a category in itself.

If an inductively-applied code does not fit under an already existing code, we discuss adding a child or parent code to the codebook. One way to assess whether an open code will make a good focused code to add is to ask whether coding additional instances of it will be helpful in the analytical stage: *will we do a code search to pull excerpts on this particular theme? Will we want to see where this particular code gets double-coded with other codes, to see what themes it most strongly co-occurs with (e.g. 'violence' and 'partner' vs 'peer' vs 'police')?* Being able to answer a question about future analysis presupposes that a team has already established an analytical process.

For our team, double-coding some excerpts is a viable and useful option, for example 'condoms' and 'PrEP' are often double-coded with 'perceptions of HIV prevention methods'. In this case, we can analyse 'PrEP' alone or 'Perceptions of HIV prevention methods' to capture the same applicable excerpt in our dataset, interspersed with other codes surrounding PrEP (e.g. 'challenges') or other methods of HIV prevention (e.g. both condoms and PrEP). The final codebook may therefore have two sets of broad codes that are intended to intersect. Depending on the coding software ultimately chosen, double coding does not add much work to the coding process. In Dedoose, for example, when a section of data is highlighted, multiple codes can be applied at the same time with a simple click. In such instances, we do caution the team to break up sections of the transcript even when a passage relates to the same general theme so that the end result, when pulling codes, is a contextualized, yet still discrete thought.

If a code already exists, we consider whether an expanded definition of the code would be helpful, and add that description to the notes or example columns, as appropriate (step seven), since "grounded theorists create codes by defining what we see in the data." (Charmaz, 2014) This is also a good time to refine the language of a code in case the existing name does not match the language the coders used in the comments. Code names should be easily found or elicited as team members code the rest of a much larger data set individually.

Steps Seven and Eight: Refine Definitions to Organise the Codes, and Finalise the Codebook

Once the codebook is fully developed, we again review definitions with the team to finalise our decisions going forward with coding the rest of the data set; for this reason, we link step seven, which could be part of step six's refinement of focused codes, with step eight. If we have incorporated *in vivo* codes into the codebook, then we adopt the definitions which participants gave to an action or incident, but we integrate them into any emergent theoretical perspectives which steps five and six offered. We also take care to streamline the codebook if necessary.

Creating the final codebook can happen at the same time as the team discusses their findings from the inductive coding exercise, or select members of the team can integrate the *a priori*

and inductive codes together afterwards, taking account of the full-team discussions. While organising codes into categories (parent codes) and sub-categories (child codes) is part of the process of axial coding, our purpose here is to give final coherence to the initial and focused codes. In our case, two members integrate the inductive findings within an *a priori* codebook using the screen share function on a video conferencing platform. Simply open and share either the desktop or two windows: the *a priori* codebook and the final merged comments document. Obviously, one could also use step six to create a codebook *de novo*, using inductive and focused codes exclusively, as in a grounded theory approach. Below, we discuss integrating an *a priori* codebook with the inductive codes from teammate comments.

We keep our *a priori* codebook in an Excel workbook, where we have separate sheets to track the interview guides from which we pulled *a priori* codes, a page for sorting codes derived from that process, the *a priori* codebook itself, another page used for integrating inductive codes into the *a priori* codebook, and a sheet with the full, or finalized, codebook. The full codebook is especially useful when more than one interview type – for example patients and providers – uses the same codebook in Dedoose (Figure 4).

During the process of integrating inductive codes into the *a priori* codebook, two team members may discover some outstanding questions which call for input from the rest of the team. These questions can be marked or highlighted, and another round of discussion, this time of the codebook itself ensues. In these cases, we send the revised codebook back to the team for discussion. Once settled, the team can pilot the resulting integrated codebook with another transcript or set of transcripts using whatever coding software the team will use for the bulk of coding.

The timing for each step of our method will depend on the size of the team and the length of the transcripts reviewed. While the main portion of the work involved a full team discussion, other steps, such as merging the documents together and preparing spreadsheets for screen-sharing, were more efficiently accomplished by a single team member.

Discussion

Main Findings

In our development of an inclusive approach towards multi-regional team-based analysis of qualitative data, we found that an iterative process using pre-existing tools can efficiently allow full team input and participation in the development of a coding framework for qualitative data sets. In the process, we addressed not only the desire from quantitative team-mates for faster returns of qualitative data, but crucially solved the recurring problem in global health partnerships of in-office versus in-the-field inclusion of teammates during data interpretation. The inclusion of on-the-ground knowledge in codebook development increased the relevance not only of what ultimately got coded, but also the resulting products which drew on those codes during the subsequent analysis phase. This had the dual benefit of increasing the validity of our findings while building capacity on the ground and orienting the research to the communities involved.

We have offered here an approach that supports global health equity through the inclusion of local expertise in the interpretation of qualitative data, because it provides a quick way to assemble individual inductive coding in geographically spread-out teams. This not only facilitates both discussion (especially among inter-disciplinary social science teams) and inter-coder comparison, facilitating the ability of teams to arrive at interpretive consensus. It also breaks through the geographical barriers by giving each coder a chance to share and contribute to the process fairly and justly. In return, this builds the confidence of the various coders since their contributions are incorporated after in-depth discussion among team members to understand their points of view. It also mitigates against the previous ‘extractive’ model in global health research where data is gathered in the global south and interpreted in the global north. Our model not only avoids that, because team members from the global south identify codes relevant to their settings which may have been completely missed or ignored by team members from the global north, but it also pays dividends by increasing the rigour of our findings and inspiring publications led by global south team members that have relevance and reach in the involved communities.

Comparison to Previous Approach

Using the process we’ve outlined above has helped us to speed up inductive coding as a team. Previously, it took long, internationally-organized, days-long workshops to combine and even to share the team’s initial line-by-line coding. With the help of the process we have outlined, we proceed both remotely and much faster – and with the benefit of easy documentation. The documentation allows us to display visually our inter-coder reliability, that is consensus across coders in their interpretations of segments of text, in ways which kappa values cannot capture. At the same time, this visual display of merged comments builds greater team consensus and confidence in one another’s perspectives, while opening up conversations which otherwise would not arise when working individually. Finally, our approach has also fostered a streamlined integration of inductive findings with *a priori* research questions in a systematic, replicable manner.

Implications of this Method

The process we have outlined saves time because every coder can work in the comfort of their geographical area without having to travel to a central place in order to discuss the inductive codes. In our view, in-person workshops with handwritten codes, which call for travel and physical meetings, is much more expensive in terms of both time and the cost of hiring a venue to meet in. (Especially during the COVID-19 pandemic, the venue needed to be spacious enough to accommodate everyone to observe social distancing requirements.) Travelling to the central place also consumed time that the team would otherwise use for discussing the codes or conducting the initial coding.

Finally, this process improved the quality of our coding because every team member had enough time to go through the transcript being discussed and give their interpretation and viewpoints. During the discussion, the team members act as the reviewers whereby each coder has the opportunity to see how other coders code a given line, and importantly, the kind of wording they used. This stimulated discussion among all coders where we aimed

to settle on a single code that encompassed many coders' viewpoints. The resulting code definitions also therefore became much clearer.

Limitations

Despite the benefits, our approach does come with several limitations, especially in under-resourced settings. Obviously, this approach requires software (MS Office suite of products), hardware (laptops), and a reliable Internet connection. Provisioning these items through research partnerships, however, builds on-the-ground infrastructural capacity beyond a onetime research study. In this regard, our method by itself is unable to address 'infrastructural' inequities (e.g. pay, physical resources), but does help address authorship inequities, as the entire team participates in coding and analysis. Further, we actively work with a Decolonizing Global Health Working Group initiative at an institution, which aims to address those additional inequities.

Second, the timing of iterations can range from several days to several weeks, depending on how frequently the team is able to meet, the size of the datasets involved (a set of short interviews can be analysed more quickly than a set of longer interviews), and the depth of discussion around how to arrange hierarchical (i.e., 'parent-child') codes. Finally, our process does not replace coding software for the actual coding of large datasets, although with smaller data sets this might be feasible. Regardless, our process does streamline codebook development, allowing the team to take account of emergent themes via inductive coding while sharing similar or divergent viewpoints along the way.

Conclusion

Using digital tools to streamline the line-by-line coding process not only creates opportunities for visual comparison across team members but also builds capacity, efficiency, and interpretive cultural rigour in multi-disciplinary, globally dispersed teams.

Acknowledgements

We would like to acknowledge the contribution of participants in the Sustainable East Africa Research in Community Health (SEARCH) studies conducted in communities in Uganda and Kenya, the leaders of these studies (Diane V. Havlir, Moses R. Kamya, Maya L. Petersen, Elizabeth A. Bukusi, James Ayieko) and other investigators including Jane Kabami, Elijah Kakande, Gabriel Chamie, Edwin D. Charlebois, Monica Getahun, Zachary Kwenya, and Catherine Koss. We appreciate the support of Kenya and Uganda Ministries of Health, and our colleagues in the Infectious Diseases Research Collaboration (IDRC) and the Kenya Medical Research Institute (KEMRI).

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Support for this work was provided by the U.S. National Institutes of Health under grants K24MH126808, (Camlin), UH3HD096915 (Havlir) and U01AI150510 (Havlir).

References

Balazs CL, & Morello-Frosch R (2013). The three Rs: How community-based participatory research strengthens the rigor, relevance, and reach of science. *Environmental Justice*, 6(1), 9–16. 10.1089/env.2012.0017

- Caliandro A, & Gandini A (2016). *Qualitative research in digital environments: A research toolkit*. Routledge.
- Castleberry A, & Nolen A (2018). Thematic analysis of qualitative research data: Is it as easy as it sounds? *Currents in Pharmacy Teaching and Learning*, 10(6), 807–815. 10.1016/j.cptl.2018.03.019 [PubMed: 30025784]
- Charmaz K (2014). *Constructing grounded theory* (2nd ed.). Sage Publications, Inc.
- Denzin NK (2017). Critical qualitative inquiry. *Qualitative Inquiry*, 23(1), 8–16. 10.1177/1077800416681864
- Deterding NM, & Waters MC (2021). Flexible coding of in-depth interviews: A twenty-first-century approach. *Sociological Methods and Research*, 50(2), 708–739. 10.1177/0049124118799377
- Gallagher K, & Kim I (2008). Moving towards postcolonial, digital methods in qualitative research: Contexts, cameras, and relationships. In *The methodological dilemma* (pp. 119–136). Routledge.
- Havliř DV, Balzer LB, Charlebois ED, Clark TD, Kwarisiima D, Ayieko J, Kabami J, Sang N, Liegler T, Chamie G, Camlin CS, Jain V, Kadede K, Atukunda M, Ruel T, Shade SB, Ssemmondo E, Byonanebye DM, Mwangwa F, & Petersen M (2019). HIV testing and treatment with the use of a community health approach in rural Africa. *New England Journal of Medicine*, 381(3), 219–229. 10.1056/nejmoa1809866 [PubMed: 31314966]
- Hemmler VL, Kenney AW, Langley SD, Callahan CM, Gubbins EJ, & Holder S (2022). Beyond a coefficient: An interactive process for achieving inter-rater consistency in qualitative coding. *Qualitative Research*, 22(2), 194–219. 10.1177/1468794120976072
- Holeman I, & Kane D (2020). Human-centered design for global health equity. *Information Technology for Development*, 26(3), 477–505. 10.1080/02681102.2019.1667289
- Israel BA, Schulz AJ, Parker EA, & Becker AB (1998). Review of community-based research: Assessing partnership approaches to improve public health. *Annual Review of Public Health*, 19(1), 173–202. 10.1146/annurev.publhealth.19.1.173
- Karnilowicz W, Ali L, & Phillimore J (2014). Community research within a social constructionist epistemology: Implications for “scientific rigor”. *Community Development*, 45(4), 353–367. 10.1080/15575330.2014.936479
- Knoblauch H, Flick U, & Maeder C (2005). Qualitative methods in Europe: The variety of social research. *Forum for Qualitative Social Research*, 6(34), 17169. 10.17169/fqs-6.3.3
- Leung MW, Yen IH, & Minkler M (2004). Community based participatory research: A promising approach for increasing epidemiology’s relevance in the 21st century. *International Journal of Epidemiology*, 33(3), 499–506. 10.1093/ije/dyh010 [PubMed: 15155709]
- Lipscomb M (2012). Abductive reasoning and qualitative research. *Nursing Philosophy*, 13(4), 244–256. 10.1111/j.1466-769X.2011.00532.x [PubMed: 22950728]
- Lock MJ, Walker T, & Browne J (2021). Promoting cultural rigour through critical appraisal tools in First Nations peoples’ research. *Australian and New Zealand Journal of Public Health*, 45(3), 210–211. 10.1111/1753-6405.13097 [PubMed: 33900687]
- Minkler M (2004). Ethical challenges for the “outside” researcher in community-based participatory research. *Health Education and Behavior*, 31(6), 684–697. 10.1177/1090198104269566 [PubMed: 15539542]
- Pratt B, & Hyder AA (2018). Designing research funding schemes to promote global health equity: An exploration of current practice in health systems research. *Developing World Bioethics*, 18(2), 76–90. 10.1111/dewb.12136 [PubMed: 27878976]
- Saldaña J (2016). *The coding manual for qualitative researchers* (3rd ed.). Sage Publications, Ltd.
- Salmen CR, Magerenge R, Nduny L, & Prasad S (2022). Rethinking our Rigor Mortis: Creating space for more adaptive and inclusive truth-seeking in community-based global health research in Kenya. *Global Public Health*, 17(12), 4002–4013. 10.1080/17441692.2019.1629609 [PubMed: 31206350]
- Thunberg S, & Arnell L (2022). Pioneering the use of technologies in qualitative research—A research review of the use of digital interviews. *International Journal of Social Research Methodology*, 25(6), 757–768. 10.1080/13645579.2021.1935565
- Timmermans S, & Tavory I (2012). Theory construction in qualitative research: From grounded theory to abductive analysis. *Sociological Theory*, 30(3), 167–186. 10.1177/0735275112457914

- Vaismoradi M, Turunen H, & Bondas T (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing and Health Sciences*, 15(3), 398–405. 10.1111/nhs.12048 [PubMed: 23480423]
- Vila-Henninger L, Dupuy C, Van Ingelgom V, Caprioli M, Teuber F, Pennetreau D, Bussi M, & Le Gall C (2022). Abductive coding: Theory building and qualitative (re) analysis. *Sociological Methods and Research*. 00491241211067508.
- Wiles R, Crow G, & Pain H (2011). Innovation in qualitative research methods: A narrative review. *Qualitative Research*, 11(5), 587–604. 10.1177/1468794111413227

What is Known

- Inductive coding is a mainstay of several types of qualitative research.
- Global health partnerships are critiqued for inequitable divisions of labour and lack of on-the-ground capacity building.
- Global health partnerships face calls for greater involvement of data collectors in data interpretation, often as part of a community-based participatory research orientation.
- Collaborating on inductive, line-by-line coding can be time-intensive because it requires teams to regularly convene to share and discuss findings.
- Integrating inductive codes into *a priori* codebooks can be unwieldy after line-by-line coding.

What this Paper Adds

- Tracks line-by-line coding in an easily documented fashion.
- Provides a quick way to assemble individual inductive coding in geographically dispersed teams to facilitate both discussion (especially among interdisciplinary social science teams) and inter-coder comparison.
- Establishes a systematic method to integrate inductive and *a priori* codes.
- Builds capacity by allowing various team members to take the lead on projects using existing software technology.
- Minimises biases by allowing for greater inclusivity in comments from a geographically-spread team.
- Fosters theoretical discussions among interdisciplinary social science teams.
- Streamlines integration of inductive codes into a priori codebooks for a comprehensive but accessible codebook through inter-coder comparison of language.
- Data collectors participate in data interpretation leading to greater validity, relevance, and reach of data interpretation while building on-the-ground capacity for using these tools in future projects.

Participant Group	Potential Broad Codes	Potential Child Codes	Code Notes based on Inductive Coding	Questions to discuss; Also <i>integrating ANC Inductive coding here</i>
DCP (ANC, Men)	01. Experiences of Study (including experience of being randomised, knowledge transfer)	1. Initial impressions of study	Nuance of child codes: <i>beliefs about why they were offered the study</i> ; reflections on relationship to the study, rather than on the study itself; First thoughts on being approached; Beliefs about why recruited for study (including importance of not being alone in the study); impressions of how the study was explained to them, how they were counselled or advised. <i>Look for good quotes and double code.</i>	

Figure 1. Sample columns from a priori codebook as set for integration of inductive findings.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

I: Tell me any questions you remember having at the time. What did you ask the provider about?

P: Yes, when we were enrolling into the study, I was a little suspicious, I asked, "Why do I join the study? Why am I joining the study alone or are there are other people?" then the doctor

Jason Johnson... Feeling suspicious

Jason Johnson Peretz April 01, 2022
Questioning benefits or motives, questioning whether others are doing it too (mimetic desire?)

The image shows a text-based interaction. On the left, there are two lines of text: an 'I:' line asking for questions and a 'P:' line providing an answer. The 'P:' line contains several phrases highlighted in red: 'I was a little suspicious', 'Why do I join the study?', and 'Why am I joining the study alone or are there are other people?'. On the right, there is a comment box with a red border. It contains two entries: one from 'Jason Johnson...' with the text 'Feeling suspicious', and another from 'Jason Johnson Peretz' dated 'April 01, 2022' with the text 'Questioning benefits or motives, questioning whether others are doing it too (mimetic desire?)'. Red dashed lines connect the highlighted text in the 'P:' line to the comment box.

Figure 2.
Using the comment function to highlight text and insert a code-comment.

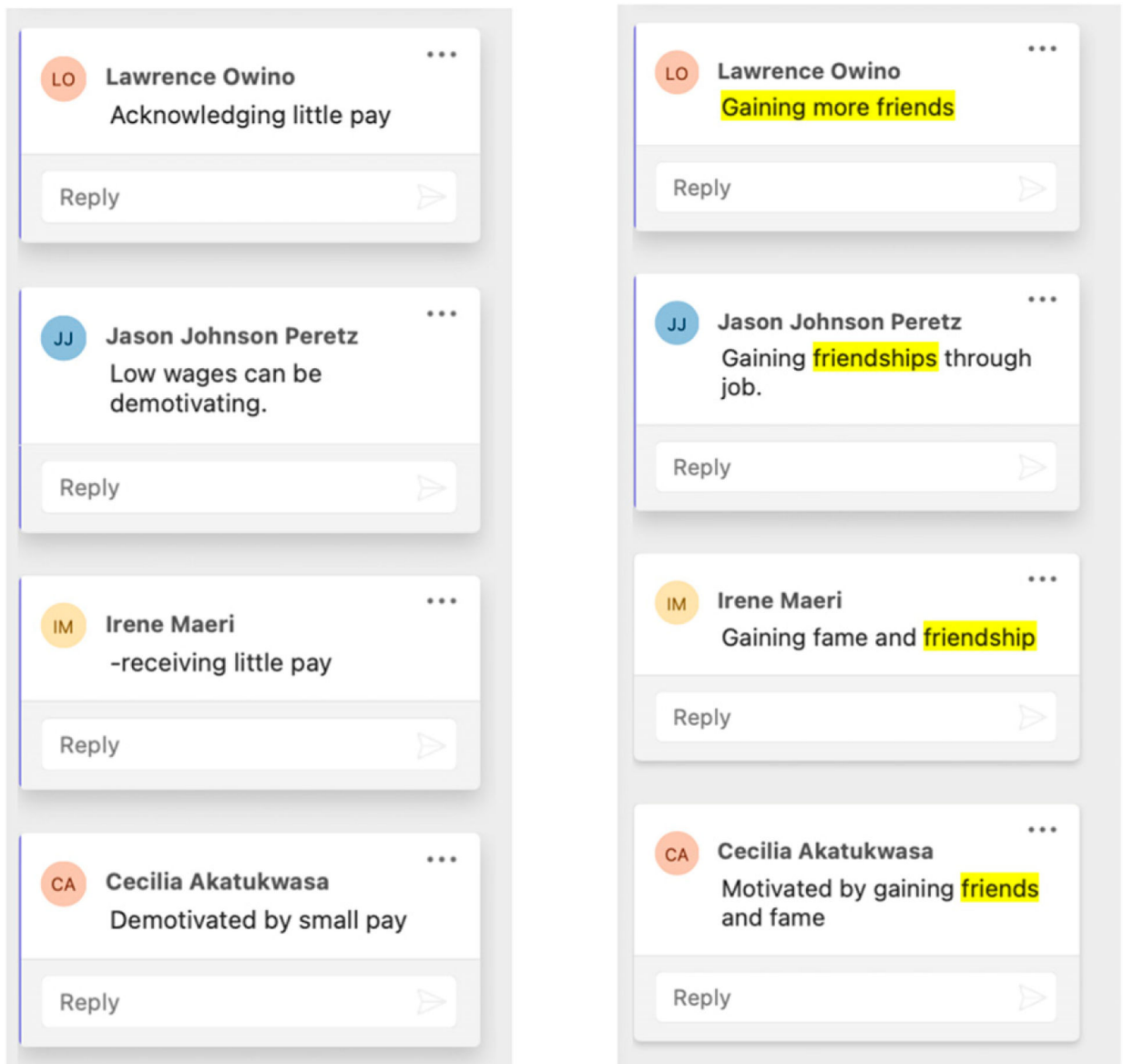
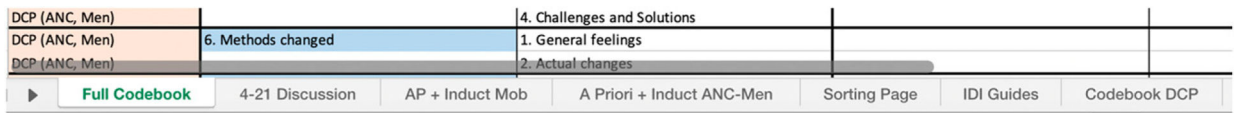


Figure 3. Two examples of comments grouped together as they appear after merging documents; the second highlights commonalities when reviewing as a team.



The image shows a portion of a spreadsheet codebook. The top part of the spreadsheet contains three rows of data. The first row has 'DCP (ANC, Men)' in the first column and '4. Challenges and Solutions' in the second column. The second row has 'DCP (ANC, Men)' in the first column, '6. Methods changed' in the second column, and '1. General feelings' in the third column. The third row has 'DCP (ANC, Men)' in the first column and '2. Actual changes' in the second column. Below the spreadsheet, there is a row of tabs. The first tab is 'Full Codebook' and is highlighted in green. The other tabs are '4-21 Discussion', 'AP + Induct Mob', 'A Priori + Induct ANC-Men', 'Sorting Page', 'IDI Guides', and 'Codebook DCP'.

DCP (ANC, Men)		4. Challenges and Solutions				
DCP (ANC, Men)	6. Methods changed	1. General feelings				
DCP (ANC, Men)		2. Actual changes				

▶ Full Codebook 4-21 Discussion AP + Induct Mob A Priori + Induct ANC-Men Sorting Page IDI Guides Codebook DCP

Figure 4.

Example of a portion of the codebook with tabs at the bottom of the spreadsheet showing the range of pages used in one workbook: The fully integrated codebook, a record of discussions, merging *a priori* and inductive codes, a page for sorting through codes to streamline them, and a page with IDI guides from which the *a priori* codebook was derived.