15. Ethnographic methods in cross-cultural research

Roseanne C. Schuster, Amber Wutich, Alexandra Brewis and Cindi SturtzSreetharan

INTRODUCTION

Ethnography is a research method rooted in efforts to better theorise people's lives in cultural context. Historically, that community of study may be based in one physical place (e.g., a village or urban neighbourhood), but increasingly it can also be defined by its intensity of social connection (e.g., transnational or virtual communities). Cross-cultural ethnographic research, by comparison, focuses on data collected across sites or from multiple groups within one site, often testing theories of cultural variability (Matsumoto & Yoo, 2006; Ember, Ember, & Peregrine 2015). Primary data collection in cross-cultural ethnography typically draws on standard ethnographic methods such as participant observation and open-ended interviews, which can be costly, time-consuming, and logistically challenging when implemented across multiple sites (Wutich & Brewis, 2019). Secondary data, such as the collections housed at the Human Relations Area Files (HRAF) Collection of Ethnography (now online as eHRAF; Levinson and Malone, 1980; Ember & Ember, 2009) or the Platform for Experimental Collaborative Ethnography (Fortun, Fortun, Bigras, Saheb, Costelloe-Kuehn, Crowder, Price, & Kenner, 2014), also allow for cross-cultural analysis with previously collected ethnographic data.

In this chapter, we outline some methodological approaches for cross-cultural ethnographic data collection and analysis of the resulting data, favouring methods that reduce the costs of data collection and burden on researchers. The methods used by ethnographers to characterise the human experience from an insider's perspective are many and complicated, but they are generally well described (Bernard, 2017). Here, we focus on cross-cultural ethnographic research designed to enable within-site thematic analysis and cross-site metatheme analysis (Wutich, Beresford, SturtzSreetharan, Brewis, Trainer, & Hardin, 2021). We cover some basics of ethnographic study design, including site selection and sampling. We explore how different approaches to data collection, including participant observation, interviewing, and group discussions, can be adapted for thematic and metathematic analysis, and discuss protocol development and analytic planning informed by cross-cultural considerations. Such approaches enable ethnographers to collect and compare novel data with the explicit purpose of being able to conduct comparisons across sites or across groups within one site. We then conclude with some thoughts on further direc-

tions and needed innovations to support cross-cultural qualitative research where comparison is analytically useful.

METHODOLOGY FOR COMPARATIVE THEMATIC ANALYSIS

Study Design

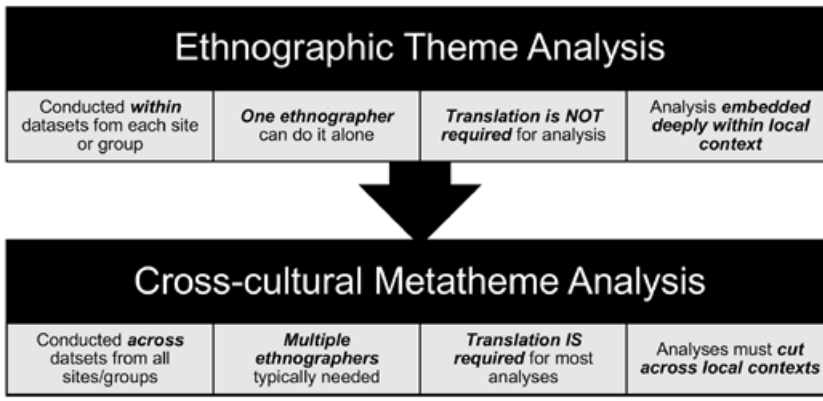
To understand how best to plan for, collect, and organise qualitative data in ways that facilitate comparisons across sites, we need to begin by recognising a model for how ethnographic data can be analysed using an ostensibly qualitative framework of thematic analysis. There are other means to achieve similar goals, but here we discuss the one that we have found more manageable to implement. This example then helps explain how procedural decisions are made at earlier stages of the project design.

Thematic analysis, as typically conducted, is well established for identifying key ideas and patterning within a site/community. This may include identifying themes that are consistently expressed within a site, or sub-themes that may be expressed by some sub-groups and not others. When investing in the effort of conducting ethnographic research within several sites/groups, it is also likely that a goal will be to analyse some aspects of the dataset across these sites/groups. A methodological strategy that anticipates both theme (within-site) and metatheme (cross-site) approaches to analysis will achieve both objectives.

The other reason we select this example is that the idea of themes will be fully familiar to most qualitative researchers. Metatheme analysis is an extension that takes an additional analytic step to identify themes shared not just within one site but across sites (Figure 15.1). The strategy outlined here, as a step-by-step guide to what to consider in designing projects for cross-site ethnographic analysis, is a purposeful one that presages both theme (within-site/group) and metatheme (cross-site/group) analysis. A metatheme can be defined as a shared meaning that occurs across contexts and expressions. They are sometimes considered broader themes in comparison to contextual group or site-specific themes (Guest, Bunce, & Johnson, 2006). For example, in a 19-site study on the effects of water insecurity on infant feeding, we identified 26 broad metathemes such as "delay in feeding", "caregiver efficacy", and "breastfeeding as a replacement for non-breastmilk foods" (Schuster, Butler, Wutich, Miller, Young, & HWISE-RCN, 2020). These broad metathemes allowed for a rich discussion of meaning within each but did not include all site-specific themes.

Site Selection

When selecting research sites, the first consideration is which sites or groups add value in terms of addressing the research question, and why. While a minimum of two sites or groups is strictly all that is needed to make a comparison, a two-site comparison is generally not recommended (Ember et al., 2015: 571), because it is nearly



Source: Modified from Wutich et al., 2021.

Figure 15.1 Overview of theme and metatheme analysis in ethnographic cross-cultural research

impossible to identify the source of any differences observed or to make meaningful attributions.

For a consequential comparison, a 2×2 design based on two related variables is the minimum number of sites or groups normally considered (Munroe & Munroe, 1991). We (see Wutich & Brewis, 2019) have reliably employed this approach numerous times since 2006 in the annual Global Ethnohydrology Study, for example sampling four sites using a 2×2 design based upon theoretically derived variables of high/low water insecurity and high/low economic development to answer questions related to perceptions of water solutions (Wutich, White, White, Larson, Brewis, & Roberts, 2014).

Another approach is to consider one theoretically driven, independently assessed criteria across several sites (typically 10 or more), as we have done in 10 countries related to expressed fat stigma (e.g., Brewis, Wutich, Falletta-Cowden, & Rodriguez-Soto, 2011). This approach can similarly be used when exploring cross-cultural understanding among multiple groups in one geographic area. For example, we (Schuster) are currently conducting analyses to identify commonalities and key differences among women experiencing pregnancy, childbirth, and postpartum during the COVID-19 pandemic, based on thematic comparison of data from nine groups representing majority and minority ethnolinguistic groups whose communities were highly impacted by COVID-19 in the Southwestern U.S.

Understanding the implications of site-level sample timing is important to accurately answer your research question (Ember et al., 2015). If conducting research related to infant and young child dietary diversity in agricultural societies, this may require sampling in different calendar months to create comparable data representing dry or "hungry" seasons. This could be similar to sampling in different calendar

months to research how specific cultures experienced the second or third wave of the COVID-19 pandemic. But, if the research question centres on the fatigue involved in adhering to social distancing or masking restrictions during COVID-19, the researcher may want to sample all groups or sites in the same calendar months even if waves of the COVID-19 pandemic occurred during different months.

Sampling

Importantly, the selection of participants must be done in a similar way across sites to yield comparable information across sites or groups (Hagaman & Wutich, 2017). Purposive sampling is the most commonly used approach for both selection of sites and participants (Liamputtong, 2020). Individual participants can be sampled from public or private spaces purposively based on certain characteristics such as gender and age, from a public space (e.g., Brewis, Wutich, du Bray, Maupin, Schuster, & Gervais, 2019), or other qualifying characteristics such as knowledge about household water management, from door-to-door surveys (e.g., Schuster et al., 2020). Snowball sampling, or recruiting from participants' networks, is preferred for "hard-to-reach" populations where other sampling techniques may fail to include those with the requisite knowledge, or where a high degree of trust may be necessary to access that group (Liamputtong, 2010, 2020).

Choosing participants at random from a complete sampling frame ("random sampling") provides a basis for later arguing the generalisability of results if that is a research goal. However, it is used infrequently in ethnographic research because its purpose is rarely to generalise to a broader population, and interviewing enough participants for a representative number is often cost-prohibitive (Ember et al., 2015). Random sampling in cross-cultural ethnographic research may, however, lend itself well to sampling homogeneous sub-groups on a very specific, shared experience. For example, we have sampled from language-specific, randomly ordered lists of all women who delivered live infants between May and December 2020 at a hospital system in Southwestern U.S. In this case, it is possible to say something about these groups undergoing the shared experience of pregnancy, birth, and postpartum during the COVID-19 pandemic, albeit only at one hospital system (though one that provides navigation services to speakers of languages other than English). If all women who met inclusion criteria were sampled from the same list, this would be an example of simple random sampling. We sampled women by language groups, an example of stratified random sampling. If participants were selected in proportion to their group's presence, this would have been an example of proportionate stratified sampling. We sought to overrepresent the experiences of refugee groups who self-reported speaking Arabic, Burmese, Kinyarwanda, and Swahili, and thus employed disproportionate random sampling to increase the numbers of women from these language groups.

Exactly how many participants constitute "the right number" has been debated for qualitative research, particularly as it uses non-probability (e.g., non-random) sampling. Drawing on its roots in theoretical saturation in grounded theory research

(Bryant & Charmaz, 2007), thematic saturation is widely applied as the point at which no new themes are identified and collecting further data would be redundant (Kerr, Nixon, & Wild, 2010). Much of current guidance pertains to individual and group interviews, though we posit similar principles may be relevant to participant observation. Identifying saturation is thus ideally an iterative process that requires an ongoing review of the data. However, this is problematic as often the sample sizes are determined before the study begins and may be limited due to funds and logistics (Guest, Namey, & McKenna, 2017). Luckily, building on a rich tradition of methodological scholarship (see Guest et al., 2017 for an overview), qualitative scholars have been working to operationalise and empirically test saturation to create guidelines for sample size estimations, in terms of saturation of both code and meaning (Guest et al., 2017; Hagaman & Wutich, 2017; Hennick, Kaiser, & Marconi, 2017; Hennick, Kaiser, & Weber, 2019). Here, we present the guidelines by criteria such as one's objective (e.g., saturation of themes or metathemes) and the chosen method of data collection (e.g., individual interviews or group discussions).

Thematic saturation (within-site) with individual interview data

Using an approach focused on completeness of the codebook in two distinct sites, Guest et al. (2006) identified 12 as the minimum number of individual interviews to capture nearly all themes within one group; after 12 interviews, new themes emerged infrequently and tend to be variations on previous themes (aligned with code saturation). Using the premise that it takes three occurrences of a theme to fully develop its meaning (Morse, Barrett, Mayan, Olson, & Spiers, 2002), Hagaman and Wutich (2017) found that three repetitions of the most common themes in unique interviews typically appeared within 16 interviews per site, concluding that their results largely supported Guest et al.'s (2006). Taking a similar approach, Hennick et al. (2017) found that 16–24 interviews were necessary to reach "saturation of meaning". Reflecting on this literature, and that a minimum of 12 interviews does not leave room for error introduced by emergent situational challenges, we suggest planning to conduct 16 interviews based on the resources available to the researcher to achieve saturation of themes within sites.

Thematic saturation (within-site) with group interview data

In a methodological study of 40 focus groups, Guest et al. (2017) found that three focus groups were sufficient to identify the most prevalent themes. In total, 80% of themes were present in two to three focus groups, and 90% of themes were present in three to six focus groups. Similarly, Hennick et al. (2019) found that four focus groups were sufficient to identify saturation of codes, but five focus groups were necessary to fully develop meaning saturation for codes that were both frequently discussed (e.g., by 51 individual participants) and infrequently discussed (e.g., by 9 participants). The literature is less developed for unstructured group interviews more common in ethnography, and so these may serve as helpful guidelines when planning thematic analyses for various group interviews.

Metathematic saturation (cross-site)

Guest et al. (2006) identified that the four metathemes in their two-group study appeared once within the first six interviews in each group. In contrast, Hagaman and Wutich (2017) found that, for the 9 metathemes in their four-site study, the first identification occurred within an average of 14 interviews, and the third identification occurred within an average of 29 interviews. Again, using the benchmark that three occurrences are necessary to develop a theme (Morse et al., 2002), they concluded that 20 to 40 interviews were needed to reach data saturation for metathemes across sites. Although there is justification for as few as six interviews per group to identify metathemes and as many as 40 to thoroughly develop meaning, we encourage those seeking to sufficiently develop the meaning of each metatheme to plan for a minimum of 16 interviews to achieve appropriate numbers for saturation of both cross-site/group metathemes and site/group-specific themes (Wutich et al., 2021). To date, we are unaware of empirically derived guidance on the minimum number of focus groups – or unstructured group interviews – to identify cross-cultural metathemes.

There are several considerations around sampling to consider. First, the guidelines we have presented for achieving saturation were developed from study populations who were homogeneous and were only sampled on a focused topic. It may take many more interviews to capture heterogeneous knowledge or opinions (Hagaman & Wutich, 2007: 35). Second, the structure of the coding system and the identification of themes also impact sample requirements (see these sections later in the chapter). That is, a researcher could use only five codes to identify broader themes, or could have 20 or 100 or many more codes to assess narrowly defined themes; this would influence how the researcher would define "saturation" in their specific dataset (Guest et al., 2006; Morse, 1995). Finally, the complexity of the topic and themes discussed may play a role; more individual interviews or group discussions may be needed with increased complexity or abstract nature of topics of interest (Guest et al., 2017).

Data Collection

Data collection in ethnography happens through a wide range of methods, both qualitative and quantitative (Bernard, 2017). Here, we focus on those methods ethnographers most often use to collect qualitative data as a prelude to thematic analysis: participant observation, ethnographic interviews, and group discussions (see Sangaramoorthy & Kroeger, 2020 for a recent wider discussion of available methods).

Participant observation

Participant observation is a hallmark of ethnographic research. Sometimes called "deep hanging out" (Geertz, 1998), participant observation involves gaining entry to a group, watching others as they navigate social situations and enact cultural rituals, learning to behave as insiders do, and using language that is understandable and sensitive to building relationships with the community (Liamputtong, 2020).

There is a large literature on working with communities, developing rapport, selecting informants, and taking into account positionality and the need for reflexivity (Bernard, 2017). These considerations are equally important in single-site/group and cross-cultural analysis, but become much more complex whether one ethnographer is navigating multiple sites/groups (e.g., Mendenhall, 2019) or multiple ethnographers are working in different sites or with different groups (e.g., SturtzSreetharan, Brewis, Hardin, Trainer, & Wutich, 2021).

Ethnographic interviews

With the goal of understanding cultural knowledge, norms, and variability, ethnographic interviewing is meant to elicit something beyond (or in addition to) people's lived experiences. Sometimes operationalised as "cultural domain interviews", interviews can elicit perspectives and understandings on how people's cultural understandings of reality are structured (Bernard, 2017). However, ethnographers tend also to intentionally elicit relational data—including but not limited to data on kinship, social networks, and semantic networks (Schnegg & Lowe, 2020). Ethnographic interviews often put relatively more emphasis on language use (oral, signed, and non-verbal) than other interviewing approaches. This presents an enormous challenge for cross-cultural designs, as researchers in every site need to be skilled in eliciting, interpreting, and analysing these data in local terminology, using local metaphors, and so on.

Group interviews

Group interviews are common in ethnographic research (Bernard, 2017; Sangaramoorthy & Kroeger, 2020). Often, they occur informally during participant observation—for example, as a group of women cook together in a shared kitchen and narrate how to prepare a dish. Ethnographers collect group interview data in more structured settings too, such as focus groups. Group interviewing can be guided by methods documented in the focus group literature (Morgan 1996; Liamputtong, 2011, 2020), as well as literatures on participatory action, community-based participatory research, photovoice, future scenarios, and other group-based methods (e.g., Roque, Wutich, Quimby, Porter, Zheng, Hossain, & Brewis, 2022). Participants in focus groups, for example, build upon each other's responses to generate new information (Kitzinger, 1995), with complex group dynamics influencing this process so that each focus group is considered a single observation (Bernard, Wutich, & Ryan, 2016; Liamputtong, 2011, 2020). Group interviews can be more complex to transcribe and analyse than individual interviews even in single-site analysis (Schuster, Brewis, Wutich, Safi, Elegido Vanrespaille, Bowen, SturtzSreetharan, Ochandarena, in review), and perhaps especially so in comparative designs (Liamputtong, 2011, 2020).

Certain methodologies may be more acceptable or empowering for research participants, e.g., structured and unstructured group interviews facilitating participants to "write their own culture together" (Liamputtong, 2011) or photovoice as a means to engage children as equal research partners (Abma & Schrijver, 2020), while other

methods may be better for reaching vulnerable individuals (see Liamputtong, 2007, 2020 for a detailed discussion). Another consideration is which methods can yield more usable information on sensitive topics. Group interviews, for example, appear to better elicit sensitive health information, while individual interviews typically perform better for sensitive topics related to the environment and perform somewhat better related to justice (Schuster et al., in review).

Protocol Development

One of the most challenging facets of cross-cultural research is developing protocols that elicit comparable data while taking into consideration the languages, cultures, and social norms specific to each site or group (Wutich & Brewis, 2019). This includes ensuring functional equivalency (preserving the meaning of a concept in translation) and cultural equivalency (how one culture interprets the underlying meaning of that question) (Peña, 2007). Here, we outline points that a researcher may want to consider in their study design and budget to address equivalence.

The first and most fundamental component is the translation from the original language of the protocol/tools into one or more other languages or language varieties. A standard approach is for the protocol/tools to be translated by one qualified person, and then back-translated by another individual blinded to the original document (Sperber, 2004). (This procedure is often required by Institutional Review Boards during ethics review.) Another approach is translation by committee, where two or more individuals work on translation independently or together, and then compare versions to develop a consensus. We have used this approach with oral translations when we are working with a language that is not commonly written. In addition to these approaches, there are a host of additional important considerations including but not limited to who translates the tools (e.g., a bicultural member of the research team, a member of the research team with deep familiarity, or a professional translator who is a non-member of the research team) and how differences are resolved across versions (Liamputtong, 2010, 2020; see also Chapter 1 in this volume). If probing prompts are included in the protocol, it is critical that these probes attend to local site specificities. Pretesting is another means to address issues with translation and transcription of the interviews that might later challenge analysis (Hurst, Arulogun, Owolabi, Akinyemi, Uvere, Warth, & Ovbiagele, 2015).

Cross-Cultural Coding

Often in cross-cultural research, we take a combined approach, starting with the creation of some codes based on *a priori* theory and frameworks that are central to study design and site- or group-specific knowledge the research team brings. Then we create additional codes in response to emergent patterns in the data following an initial review of the dataset and throughout the coding process as necessary. Another choice is how narrow or broad the codes are defined (Ember & Ember, 2009). Broad codes can facilitate the identification of themes within the code. Alternatively, codes

can be quite narrowly defined, so that the research uses multiple codes as a building block to generate themes. In the middle of this continuum, codes can represent themes in themselves.

Using a codebook preserves the systematic application of codes to the data. For each code, the researcher creates a definition, criteria for inclusion and exclusion, and typical, atypical, and "close-but-no" examples to set the bounds of the code (MacQueen, McLellan, Kay, & Milstein, 1998). The codebook is typically refined throughout the data analysis process as new expressions are encountered; the research team may decide to do a significant revision after a subset of data has been reviewed (e.g., after N sites/groups, or the Nth unit within all groups/sites) or periodically (e.g., every Nth unit). The entry for each code should be broad enough to encompass the various manifestations of that code in each site; noting site-specific examples as typical or atypical examples helps guide the coding. Again, combinations of approaches can be used here. For example, in the 19-site study of water insecurity and infant feeding, the codebook was significantly refined after a preliminary analysis of the first nine sites, with smaller modifications made continuously afterwards (Schuster et al., 2020). Importantly, when *significant* changes to the codebook are made such as to modify boundaries of inclusion and exclusion, the dataset is then re-reviewed.

While codebooks are important for use even by one researcher, they are a key to standardising *interrater reliability* when coding using a team-based approach that is more likely in cross-cultural studies (McHugh, 2012). For example, we commonly train student researchers on qualitative methods through hands-on coding of cross-cultural datasets in teams ranging from 2–20 individuals, making it very important to ensure systematic application of codes to the dataset (see Wutich & Brewis, 2019; Beresford, Wutich, DuBray, Ruth, Stotts, SturtzSreetharan, & Brewis, 2022). To do this, first a subset of texts collected from each research site in the study is coded collectively by the core research team, leveraging the team's cross-cultural theoretical, subject, and context-specific expertise. Then each coder is asked to review the same data subset (uncoded) and make categorical decisions of "code present" or "code absent" for each unit of analysis. (This is most clearly done when units of analysis are specified, such as in content analysis, as a tagging approach is more difficult to present as units for decisions of present or absent.) The interrater reliability (agreement of each coder with the lab standard) is then assessed using a measure of observer agreement for categorical data, developed to correct for agreement by chance, such as the kappa statistic (Landis & Koch, 1997; McHugh, 2012). Each member of the coding team repeats this process until they reach the desired level of kappa established *a priori*. Achieving a kappa above 0.8 demonstrates ideal agreement and corresponds to a minimum of 64% of data being reliably coded, although a kappa above 0.7 still represents good agreement beyond chance (McHugh, 2012). This can be challenging if there is a wide range of expressions for that code and few occurrences, such that the core team may exhaust the dataset in identifying training material. Of course, agreement can be assessed when two or more coders simultaneously apply codes, an approach fit for smaller coding teams of

ethnographers bringing the requisite theoretical, content, and contextual knowledge. Overall, our team finds this is an efficient process to train new researchers in qualitative methods, engage multiple coders in coding small or large datasets in a timely manner, and ensure high-quality coding overall (Ruth, Brewis, Blasco, & Wutich, 2019; Wutich & Brewis, 2019). Along with detailing other methodological decisions made, reporting the kappa coefficient reached as per each code for team-based coding in publications (e.g., Wutich et al., 2014) is important in establishing this as part of systematic rigour.

Identifying Themes, Then Metathemes

Themes can be derived inductively, meaning the researcher looks only to the data for the identification of themes. This is more common when the research is exploring a completely new phenomenon where little is known, and it is used commonly in grounded theory research. And themes can be derived deductively, meaning the researcher brings and applies their prior knowledge of the topic, theory, frameworks, and the sites/groups, to the identification of themes. Theme identification exists on a continuum of less structured to more structured, so that a combination of these approaches may be appropriate in many contexts (Ryan & Bernard, 2003). For example, in the study of water insecurity and infant feeding, we deductively identified four domains important to infant feeding (e.g., breastfeeding, non-breastmilk feeding, caregiver capabilities, infant health) and then combined these with inductive, cross-cultural metatheme identification and description to generate 29 metathemes (Schuster et al., 2020).

Once deciding on the inductive, deductive, or combination approach, the most straightforward and widely used criterion for analysis of codes is repetition, or the frequency with which expressions of that theme are made by multiple participants (Ryan & Bernard, 2003). Complementary criteria include universality, or how widely shared a phenomenon is across participants of certain groups/sites; and differentiation, or the relative importance of certain themes in specific groups/sites (e.g., Boateng & Adams, 2016; Schuster et al., 2020). The use of local terms, metaphors and analogies, linguistic connectors, transitions such as shifts in tone, missing data (see Ryan & Bernard, 2003), and emphasis, or emotional speech (Boateng & Adams, 2016) are additional and very important criteria in the identification of culturally embedded themes. These somewhat more nuanced techniques elicit meaning and facilitate the identification of themes that could be overlooked without rich ethnographic context.

In writing up the results of thematic analysis, it is standard practice to provide thick description and ethnographic exemplars. While ethnographic writing encompasses a range of genres and experiments (Atkinson, 2019), we cover two easy-to-use techniques in our discussion and illustrate how each can be used for thematic and metathematic analyses: thick description (below) and ethnographic exemplars (at section end). These writing techniques enable ethnographers to explain and illustrate

the ways in which conclusions were drawn. They are also sensible means by which to proceed with metatheme analysis.

Thick description is a form of analysis that presents the fine details of social moments, and interprets them in ways that draw out social and cultural meanings of broader significance (Geertz, 1973: 318). Thick description tends to provide rich and vivid images of the phenomenon being described, in ways that are meant to be striking and illuminating for the reader. In cross-cultural research, it can be particularly challenging to provide thick descriptions from across multiple sites in ways that are analytically coherent. The cross-cultural ethnography *Fat in Four Cultures* examines weight and body perceptions across four sites that varied in openly expressed stigma about body size and average body weight, and presents within-site thick descriptions in separate chapters (SturtzSreethran et al., 2021). In each, ethnographers richly describe experiences from their research sites, and later culminate in a presentation of metathemes based on this ethnography.

In an excerpt from her cross-cultural ethnographic work in Japan, for example, Cindi SturtzSreetharan writes (pp. 39–40),

> On a sweaty summer day in the late 1980s, Cindi went on an outing with her friend, Yōko, and Yōko's family. They were getting into the family's car for a day trip to see Shinto shrines in the old Japanese city of Kyoto. Yōko's father pointed at his daughter's legs. "Oi! Daikon ashi da na!" he declared. "Hey! You have daikon legs!" Daikon, the long tapering white radish, wide at the top, but narrow at the bottom, is found in many pickled condiments and meals on Japanese tables. Even though Cindi was still learning Japanese at the time, she instantly recognized that having one's legs compared to a thick root vegetable was not a good thing. The closest American English equivalent would be telling someone that they have "cankles" ("calf" and "ankle" blended to cankle). At the time, nobody replied to Yōko's father's comment. His less-than-flattering statement hung in the air, unaddressed.
>
> Decades later, now fluent in Japanese, Cindi was again in Osaka, this time doing the research for [the *Fat in Four Cultures*] book. A different friend, Reiko, made a notable confession. Several weeks before, Cindi had interviewed her about fat, health, and fitness as they play out in everyday life in Japan. In responding to the interview questions about whether she noticed large-bodied people, Reiko had been quite adamant she didn't really notice "futotteru" (fat) people and that she certainly didn't have anything against them. At that moment, they were driving past a large man who was standing in the road directing traffic. Dressed in navy blue overalls and boots, the overalls made his stomach protrude in an obvious way. This seemed to jog Reiko's memory. She began to recount a story in detail about a typically hot and humid summer day at her work in an educational administration office. She had looked up from her computer to find a different man standing in front of her desk. Her office was pleasantly air conditioned, and the hallway he had been waiting in was not, so he seemed to be there simply to enjoy the cooler air. She described him as a large man—larger, she noted, than the road worker they had just passed. "He was dripping with sweat and just standing there," she said in disgust-tinged recollection. She then admitted that what she had told Cindi in the formal interview—that futotteru people didn't bother her—actually wasn't true. That man in her office had really bothered her, and it was simply because he was futotteru. She didn't want him anywhere near her, she now remembered.
>
> These two events that took place 30 years apart in Osaka reflect a complex grappling with what it means to be fat. They tap into deep concerns Osakan people (and Japanese more generally) share about futotteru hito, "fat people." They also highlight that women and men talk differently about bodies, and that age and one's own body size also influence these

> discourses. The fundamental theme across all conversations, however, is that in Japan, for most ordinary folks, being futotteru is a very bad thing.[1]

As this example illustrates, reflexivity and positionality are core concerns for ethnographers. In thick description, it is important to *show* the reader important ethnographic details, not just tell them. This thick description led to the generation of themes such as *fat is linked to disgusting sweat*. Reviewed in context with themes generated by additional thick descriptions, this theme found in the Japan site contributed to the building of a *fat is a social failure* metatheme found across sites.

Much less is written about the process to conduct *metathematic analysis*, but it is based in many of the same principles as thematic analysis and the goal of searching for either inductive or deducting patterning. Wutich and colleagues (2021) detail the team-based process used to identify metathemes in *Fat in Four Cultures*. First, the ethnographers who completed data collection conducted an initial thematic analysis *within* each of the sites using a variety of theme identification techniques in the language of data collection (Ryan & Bernard, 2003). Then these same ethnographers engaged in a collaborative process where the site-specific themes were combined in a common language. By working together, the ethnographers then sorted the complete list of site-specific themes into metathemes, using such techniques as pile sorting and group discussion, and reaching consensus through dynamic conversation. (Of note, to engage in this process, all the sites should be organised and evaluated similarly for comparability—see prior sections on Site Selection, Sampling, and Protocol Development.) Metatheme analysis can similarly engage students in coding as described in the team coding section; in this case, each site's lead ethnographer closely collaborates on the development of the codebook and then analysis of the coded data to identify site-specific themes, and works with the other lead ethnographers as a team to build metathemes as described here.

Similarly, both thick descriptions and ethnographic exemplars as developed for purposes of site-specific thematic analysis can be built upon to yield metathemes. We have provided an example of how thick description serves as the basis to generate themes and build to metathemes. Ethnographic exemplars are ethnographic observations that are richly described, unpacked, and analysed as representative of larger cultural trends. Ethnographers can use direct quote, slice-of-life description, synecdoche, or ethnographic moment as the basis for ethnographic exemplars, which can provide building blocks for metatheme analysis. Following their presentation, ethnographers can draw out cross-cultural comparisons, richly illustrating similarities and differences across sites.

Here, we provide another example from the cross-cultural ethnography *Fat in Four Cultures* (SturtzSreetharan et al., 2021). In this case, ethnographic exemplars are presented as direct quotes that exemplify the views of people from four cross-cultural sites on the gendered and aged nature of fat (Table 15.1). The metatheme "Fat is a problem for all ages" was identified during a pile sorting exercise wherein it was noticed that no one escaped the negative sentiment of fat. That is, as the direct quote exemplars below show, young children are discursively created

Table 15.1 *Exemplars for the metatheme "Fat is a problem for all ages," from the* Fat in Four Cultures *study*

Study Site	Exemplar Quotes
Osaka, Japan	"Men gain weight, too, after marriage and kids, but not as much [as women]. Since getting married, my husband has gained 10 kg, in the past he was much thinner, but now his belly is really quite big." (Sachiko, woman, 38 years old)
North Georgia, United States	"I can't think of a single reason [where having a bigger body is an advantage, for men or women of any age], unless it's to help the fabric industry. … Or you could sit on somebody, I guess, if somebody broke into the house." (Rosemary, woman, in her 60s)
Encarnación, Paraguay	My daughter, for example, she's nine, a girl, and she's big-boned. She's big-boned … chubby, she's got a big belly. She's fat, in general. Fat. Many people say, "She's so fat! What is she eating so much?" And I tell her, "Don't feel bad." … I've noticed that there comes a time when young people … begin to limit their food intake. (Antonia, woman, 52 years old)
Apia, Samoa	"When I returned from overseas, especially in my village and my family, and I notice that, there's something going on here. So, I've been asking myself why is it changing now. Before, back in the day, they were—they all looked strong, looked fit. I try to encourage them, but like half the community, poor people are sick." (Amosa, man, 38 years old)

as "fat", indeed, both men *and* women of various ages are pronounced "fat" across the sites. The metathemes depict specific comparisons from each set of site-specific interviews, providing a glimpse into the lived experience of the participants in each site. Having multiple analysts can facilitate the identification of engaging, illustrative ethnographic exemplars for writing up metatheme analysis in cross-cultural research (Wutich et al., 2021).

CONCLUSION AND FUTURE DIRECTIONS

While ethnographic methods have mostly been applied to conduct site- or community-specific analyses, the same methods can be reasonably applied to comparative cultural analyses as well. Here, we have identified some ways in which methods from basic and well-established modes of ethnographic data collection and then thematic analysis can be applied to this goal. The procedures we have highlighted allow for within-site/group analyses too, which recognises the investments in time and effort that any ethnographic data collection takes, and that some themes of interest may be important locally but not when compared across sites. Explicit comparison, however, using such standard ethnographic techniques as theme analysis requires significant planning to ensure comparability for analysis, such as adequate sampling and site justification, and consistent data collection sufficient to allow post-field comparisons to be made meaningfully. There are few examples of metatheme analysis, and much room for innovation in methods for detecting, comparing, and integrating metathemes and sub-metathemes. For example, the process of making decisions around metathemes could be moved from in-person discussions

to being systematised using network or similar software-based analysis (see Brewis, Wutich, Galvin, and Lachaud, 2022 for an example using word-based analysis).

Finally, globalisation shifts how ethnography can or should be done; this includes the necessity of treating culture as a dynamic phenomenon that can easily transcend physical place. Ethnographic methods are being tested and refined to meet this recognition, as we have here with metathematic analysis. It is perhaps surprising how few such efforts are immediately evident in the literature, given a more general need to test and apply a greater ethnographic toolkit. Different approaches to cross-cultural ethnography should provide varied insights for those interested in explaining, and even harnessing the power of, culture in shaping human behaviour and the well-being of peoples, environment, institutions, and systems in our globalised world. Cross-cultural ethnography is overall a relatively underdeveloped approach, and more experimentation with comparative methods should usefully expand the ethnographic toolkit.

ACKNOWLEDGEMENTS

This work was supported by the US National Science Foundation (NSF SBE-2017491) Cultural Anthropology Methods Program: Advanced Research Training for Ph.D. Students. We acknowledge the support of the Virginia G. Piper Charitable Trust in support of the Mayo Clinic/ASU Obesity Solutions initiative.

NOTE

1. Reprinted with permission of the publisher.

REFERENCES

Abma, T. A. & Schrijver, J. (2020). "Are we famous or something?" Participatory health research with children using photovoice. *Educational Action Research*, 28(3):405–426. doi .org/10.1080/09650792.2019.1627229

Atkinson, P. (2019). *Writing ethnographically*. Walnut Creek, CA: SAGE Publications.

Beresford, M., Wutich, A., DuBray, M., Ruth, A., Stotts, R., SturtzSreetharan, C. & Brewis, A. (2022). Coding qualitative data at scale: Lessons learned for large coder teams based on 18 studies. *International Journal of Qualitative Methods*, 21:1–15. doi:10.1177/16094069221075860

Bernard, H. R. (2017). *Research methods in anthropology: Qualitative and quantitative approaches*. Lanham, MD: Rowman & Littlefield.

Bernard, H. R., Wutich, A. & Ryan, G. W. (2016). *Analyzing qualitative data: Systematic approaches* (2nd edn). Thousand Oaks, CA: SAGE Publications, Inc.

Boateng, G. O. & Adams, T. L. (2016). "Drop dead… I need your job": An exploratory study of intra-professional conflict amongst nurses in two Ontario cities. *Social Science & Medicine*, 155:35–42. https://doi.org/10.1016/j.socscimed.2016.02.045

Brewis, A., Wutich A, du Bray, M., Maupin, J., Schuster, R.C. & Gervais, M. (2019). Community hygiene norm violators are consistently stigmatized: Evidence from four global sites and implications for sanitation interventions. *Social Science & Medicine*, 220:12–21. doi:10.1016/j.socscimed.2018.10.020

Brewis, A. A., Wutich, A., Falletta-Cowden, A. & Rodriguez-Soto, I. (2011). Body norms and fat-stigma in global perspective. *Current Anthropology*, 52:269–76.

Brewis, A., Wutich, A., Galvin, M. and Lachaud, J. (2022). Localizing syndemics: A comparative study of hunger, stigma, suffering, and crime exposure in three Haitian communities. *Social Science & Medicine*, 295:113031. doi: 10.1016/j.socscimed.2020.113031.

Bryant, A. & Charmaz, K. (Eds.). (2007). *The SAGE handbook of grounded theory*. London: SAGE Publications.

Ember, C. R. & Ember, M. (2009). *Cross-cultural research methods*. Walnut Creek, CA: AltaMira.

Ember, C. R., Ember, M. & Peregrine, P. N. (2015). Cross-cultural research. In H. R. Bernard & C. C. Gravlee (Eds.), *Handbook of methods for cultural anthropology* (2nd edn), pp. 561–600. London: Rowman & Littlefield.

Fortun, K., Fortun, M., Bigras, E., Saheb, T., Costelloe-Kuehn, B., Crowder, J., Price, D. & Kenner, A. (2014). Experimental ethnography online: The asthma files. *Cultural Studies*, 28(4):632–642.

Geertz, C. (1998). Deep hanging out. *The New York Review*, 45(16):69–72.

Geertz, C. (1973/2008). *Thick description: Toward an interpretive theory of culture*. Abingdon, UK: Routledge.

Guest, G., Bunce, A. & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, 18:59–82. doi:org/10.1177/1525822X05279903

Guest, G., Namey, E. & McKenna, K. (2017). How many focus groups are enough? Building an evidence base for nonprobability sample sizes. *Field Methods*, 29(1):3–22. doi:10.1177/1525822X16639015

Hagaman, A. K. & Wutich, A. (2017). How many interviews are enough to identify metathemes in multisited and cross-cultural research? Another perspective on guest, Bunce, and Johnson's (2006) landmark study. *Field Methods*, 29(1):23–41. https://doi.org/10.1177/1525822X16640447

Hennick, M. M., Kaiser, B. N. & Marconi, V. C. (2017). Code saturation versus meaning saturation: How many interviews are enough? *Qualitative Health Research*, 27(4):591–608. doi:10.1177/1049732316665344

Hennick, M. M., Kaiser, B. N. & Weber, M. B. (2019). What influences saturation? Estimating sample sizes in focus group research. *Qualitative Health Research*, 29(10):1483–1496. doi:10.1177/1049732318821692

Hurst, S., Arulogun, O. S., Owolabi, A. O., Akinyemi, R., Uvere, E., Warth, S. & Ovbiagele, B. (2015). Pretesting qualitative data collection procedures to facilitate methodological adherence and team building in Nigeria. *International Journal of Qualitative Methods*, 14:53–64. doi:10.1177/160940691501400106

Kerr, C., Nixon, A. & Wild, D. (2010). Assessing and demonstrating data saturation in qualitative inquiry supporting patient-reported outcomes research. *Expert Review of Pharmacoeconomics & Outcomes Research*, 10:269–281. doi: 10.1586/erp.10.30.

Kitzinger, J. (1995). Qualitative research. Introducing focus groups. *BMJ*, 311:299–302. doi:10.1136/bmj.311.7000.299

Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33:159–174.
Levinson, D. & Malone, M.J. (1980). *Toward explaining human culture: A critical review of the findings of worldwide cross-cultural research*. New Haven, CT: HRAF Press.
Liamputtong, P. (2007). *Researching the vulnerable: A guide to sensitive research methods*. London, UK: SAGE Publications.
Liamputtong, P. (2010). *Performing qualitative cross-cultural research*. Cambridge, UK: Cambridge University Press.
Liamputtong, P. (2011). *Focus group methodology: Principle and practice*. London, UK: Sage Publications, Ltd.
Liamputtong, P. (2020). *Qualitative research methods* (5th edn). South Melbourne: Oxford University Press.
MacQueen, K., McLellan, M. E., Kay, K. & Milstein, B. (1998). Codebook development for team-based qualitative analysis. *Cultural Anthropology Methods*, 10:31–36. doi.org/10.1177/1525822X980100020301
Matsumoto, D. & Yoo, S. H. (2006). Toward a new generation of cross-cultural research. *Perspectives on Psychological Science*, 1(3):234–250. doi.org/10.1111/j.1745-6916.2006.00014.x
McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3):276–282.
Mendenhall, E. (2019). *Rethinking diabetes: Entanglements with trauma, poverty, and HIV*. Ithaca, NY: Cornell University Press.
Morgan, D. L. (1996). *Focus groups as qualitative research* (Vol. 16). Thousand Oaks, CA: SAGE Publications.
Morse, J. (1995). The significance of saturation. *Qualitative Health Research*, 5:147–149. doi: org/10.1177/104973239500500201
Morse, J. M., Barrett, M., Mayan, M., Olson, K. & Spiers, J. (2002). Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods*, 1:13–22. doi.org/10.1177/160940690200100202
Munroe, R. L. & Munroe, R. H. (1991). Results of comparative field studies. *Cross-Cultural Research*, 25(1–4):23–54.
Peña, E. D. 2007. Lost in translation: Methodological considerations in cross-cultural research. *Child Development*, 78(4):1255–1264. doi.org/10.1111/j.1467-8624.2007.01064.x
Roque, A., Wutich, A., Quimby, B., Porter, S., Zheng, M., Hossain, M. J. & Brewis, A. (2022). Participatory approaches in water research: A review. *WIREs Water*, 9:e1577. doi: 10.1002/wat2.1577.
Ruth, A., Brewis, A., Blasco, D. & Wutich, A. (2019). Long-term benefits of short-term research-integrated study abroad. *Journal of Studies in International Education*, 23(2):265–280. doi: 10.1177/1028315318786448
Ryan, G. W. & Bernard, H. R. (2003). Techniques to identify themes. *Field Methods*, 15(1):85–109. doi: 10.1177/1525822X02239569
Sangaramoorthy, T. & Kroeger, K. A. (2020). *Rapid ethnographic assessments: A practical approach and toolkit for collaborative community research*. London: Routledge.
Schnegg, M. & Lowe, E. D. (Eds.). (2020). *Comparing cultures: Innovations in comparative ethnography*. Cambridge, UK: Cambridge University Press.
Schuster, R. C., Brewis, Wutich, A., A., Safi, C., Elegido Vanrespaille, T., Bowen, G., SturtzSreetharan, C., McDaniel, A. & Ochandarena, P. (in review). Individual interviews versus focus groups for evaluations of international development programs: Systematic testing of method performance to elicit sensitive information in a justice study in Haiti.
Schuster, R. C., Butler, M. S., Wutich, A., Miller, J. L., Young, S. L. & HWISE-RCN. (2020). "*If there is no water, we cannot feed our children*": The far-reaching consequences of water

insecurity on infant feeding practices and infant health across 16 low- and middle-income countries. *American Journal of Human Biology*, 32:e23357. doi.org/10.1002/ajhb.23357

Sperber, A. D. (2004). Translation and validation of study instruments for cross-cultural research. *Gastroenterology*, 126:S124–S128. doi:10.1053/j.gastro.2003.10.016

SturtzSreetharan, C., Brewis, A., Hardin, J., Trainer, S. & Wutich, A. (2021). *Fat in Four Cultures: A Global Ethnography of Weight*. Toronto, ON: University of Toronto Press.

Wutich, A. & Brewis, A. (2019). Data collection in cross-cultural ethnographic research. *Field Methods*, 31(2):181–189. doi.org/10.1177/1525822X19837397

Wutich, A., Beresford, M., SturtzSreetharan, C., Brewis, A., Trainer, S. & Hardin, J. (2021). Metatheme analysis: A qualitative method for cross-cultural research. *International Journal of Qualitative Method*s, 20:1–11. doi:10.1177/16094069211019907

Wutich, A., White, A. C., White, D. D., Larson, K. L., Brewis A. & Roberts, C. (2014). Hard paths, soft paths or no paths? Cross-cultural perceptions of water solutions. *Hydrology and Earth Systems Science*, 18:109–120. doi:10.5194/hess-18-109-2014