

Why and How Epidemiologists Should Use Mixed Methods

 Lauren C. Houghton^{a,b} and  Alejandra Paniagua-Avila^a

Abstract: The field of epidemiology's current focus on causal inference follows a quantitative approach and limits research questions to those that are strictly quantifiable. How can epidemiologists study biosociocultural public health problems that they cannot easily quantify? The mixed-methods approach offers a possible solution by incorporating qualitative sociocultural factors as well as the perspective and context from the population under study into quantitative studies. After a pluralist perspective of causal inference, this article provides a guide for epidemiologists interested in applying mixed methods to their observational studies of causal identification and explanation. We begin by reviewing the current paradigms guiding quantitative, qualitative, and mixed methodologies. We then describe applications of convergent and sequential mixed-methods designs to epidemiologic concepts including confounding, mediation, effect modification, measurement, and selection bias. We provide concrete examples of how epidemiologists can use mixed methods to answer research questions of complex bio-socio-cultural health outcomes. We also include a case study of using mixed methods in an observational study design. We describe how mixed methods can enhance how epidemiologists define underlying causal structures. Our alignment of mixed-methods study designs with epidemiologic concepts addresses a major gap in current epidemiology education—how do epidemiologists systematically determine what goes into causal structures?

Submitted April 5, 2022; accepted November 11, 2022

From the ^aDepartment of Epidemiology, Columbia University Mailman School of Public Health, New York, NY; and ^bHerbert Irving Comprehensive Cancer Center, Columbia University, New York, NY.

L.H. conceived the idea for the article; developed the ideas for and prepared the sections "Applications of mixed methods to epidemiologic research process," "Example of an epidemiologic, observational study using mixed methods," "Conclusion"; and critically revised all sections of the article. A.P.-A. contributed substantially to the conceptualization of the article; prepared the sections "Introduction," "Quantitative, qualitative, mixed methods," contributed to the "Applications of mixed methods to epidemiologic research process" and "Example of epidemiologic, observational study using mixed methods" sections; conducted the literature review; and critically reviewed the final version.

L.C.H. was supported to conduct this work through National Cancer Institute K07CA218166.

The authors report no conflicts of interest.

There are no data included in this article.

Correspondence: Lauren C. Houghton, Department of Epidemiology, Columbia University Mailman School of Public Health, 722 West 168th Street, New York, NY 10032. E-mail: lh2746@cumc.columbia.edu.

Copyright © 2022 The Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the Creative Commons Attribution License 4.0 (CCBY-NC-ND), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ISSN: 1044-3983/23/342-175-185

DOI: 10.1097/EDE.0000000000001565

Keywords: Mixed methods; Qualitative; Causal inference

(*Epidemiology* 2023;34: 175–185)

Health outcomes are the product of complex social and biologic factors that interact at the molecular, individual, organizational, and broader ecologic levels over time.^{1–3} Historically, the interdisciplinary nature of epidemiology positioned epidemiologists to study health across these levels. Overtime, epidemiology has become focused on causal inference, a process that consists of contrasting health outcomes among two or more groups of participants under different exposures.⁴ Ideally, epidemiologists would approach causal inference using interdisciplinary methodologies;⁵ however, causal inference in epidemiology follows a quantitative approach⁴ and is increasingly methods driven.⁶ Epidemiologists seldom overtly use qualitative approaches drawn from anthropology and other social sciences.⁷ For this reason, multiple authors have argued that modern epidemiology limits research questions to those that are strictly quantifiable.^{7,8} As Krieger and Davey-Smith state, "Causes do not cease being causes if they are challenging to study or to address."⁹ Although the call echoes for epidemiologists to study biomedical and social causes of disease,¹⁰ it is unclear how to integrate them within one study, how to capture social constructs that are difficult to quantify such as contextual factors, and how to incorporate the population's perspective. Mixed-methods research offers solutions.

Mixed-methods research integrates quantitative and qualitative data within a single study and is similar to how epidemiologists conceive triangulation, a concept suggested as essential in improving causal inference in epidemiology from a pluralist perspective.^{9,11,12} Mixed methods can bring the population's insight into hypothesis generation and incorporate context into causal structures. Epidemiology training programs may offer limited instruction in mixed-methods research, and epidemiologists might be unsure about how to apply these methods. This article provides a guide for epidemiologists to design mixed-methods studies, with a focus on epidemiologic concepts including confounding, selection bias, attrition, measurement error, mediation, and effect modification. We see the applications being most relevant to observational studies designed for causal identification and

causal explanation.¹¹ First, we summarize the current paradigms guiding quantitative, qualitative, and mixed methodologies. Then, we describe specific applications of mixed methods to epidemiologic research. We include examples of existing and hypothetical studies to illustrate the alignment of epidemiologic concepts with mixed-methods study designs. A case study illustrates how to implement mixed methods in an observational study. The third part describes how to use mixed methods to define underlying causal structures. We conclude with current limitations of applying mixed methods to epidemiology.

QUANTITATIVE, QUALITATIVE AND MIXED-METHODS PARADIGMS AND DESIGN

Comparing Quantitative and Qualitative Methods

Qualitative and quantitative research differ in their foundational scientific paradigms.^{13,14} The quantitative paradigm, rooted in positivism and empiricism, reduces phenomena to empirical indicators that represent the truth.¹⁵ In contrast, the qualitative research paradigm is based on constructivism, in which reality is socially constructed and constantly changing.^{14,15} In terms of approach, quantitative methods are primarily deductive—they move top-down from theory, to the formulation of hypothesis, and then to confirmation or rejection by individual observations.^{14,16} Qualitative methods, in contrast, are primarily inductive—they move bottom-up from particular observations, to patterns, to the formulation of hypotheses, and then to theories.¹⁶ Furthermore, epidemiologists usually interpret quantitative data from an etic or external perspective.¹⁶ On the other hand, qualitative research adheres to the approach traditionally followed by anthropologists, characterized by an emic perspective that puts participants and their views at the center of research.^{7,17} For example, in nutritional epidemiology an etic view may use methods to obtain nutrient level data, whereas an emic view may use methods to understand cultural practices around meals.

Quantitative and qualitative data collection and analysis also differ. First, quantitative methods gather numerical data, typically from close-ended and structured questionnaires, publicly available data resources, clinical records, or biologic measurements, whereas sources of qualitative information include text and images coming from documents, transcriptions, or field notes derived from in-depth interviews, focus groups, and participant observations. Second, quantitative data collection usually occurs separately and before data analysis. In contrast, qualitative data collection tends to be more iterative.¹⁸ Qualitative researchers may refine their interview guide and analyze data, as they collect it to help assess saturation—that is, when no new themes emerge from additional participants. Third, quantitative data collection tends to be generated from a probabilistic sample with the goal of being generalizable, while qualitative data collection follows a purposeful sampling strategy to gain in-depth information.

Whether purposeful or probabilistic, both sampling strategies capture elements of similarity and differences¹⁹ and, in reality, observational studies often collect quantitative data less from probabilistic and more from convenient samples. Some researchers argue that opposing paradigms justify keeping quantitative and qualitative approaches separate²⁰; we and others argue, however, that they are complementary, as each method can access different aspects of a research problem that cannot be accessed with one method alone.¹⁹ For example, mixed-methods research can help to assess generalizability by including a large and representative sample for quantitative analysis and collecting qualitative data to gauge if the local context reflects the larger one.

Mixed-methods Paradigm and Design

The pragmatic mixed-methods paradigm¹³ prioritizes the research question over the methods used to answer it.¹⁴ Also following pragmatic ontology, other epidemiologists argue that causal reasoning based on qualitative evidence is justified. Specifically, Bannister-Tyrrell et al. argue that Russo and Williamson²¹ and Reiss's²² theories of causal inference “align with the empirical focus of epidemiology and allow for different types of evidence to evaluate causal claims, including evidence originating from qualitative research.”²³ Bannister-Tyrrell et al. see qualitative data specifically helping with mediation (mechanism of causal relations) and effect modification (the effects of context on outcomes).²³ We agree that mixed methods can aid in improving causal explanation, but we also see it improving causal identification. Causal identification includes identifying an association between an exposure and outcome and eliminating alternative explanations for that association through taking into account confounding and reducing sources of bias.²⁴ Both causal identification and explanation can be improved through implementing mixed-method design into epidemiologic studies.

Mixed-methods designs, based on Creswell's 2018 update,¹⁴ include the convergent, explanatory sequential, and exploratory sequential designs, differentiated by the order in which the methods are used, the stage at which the data are integrated, and the emphasis of each method data relative to each other. The convergent design places equal priority on both methods, by simultaneously collecting parallel qualitative and quantitative data and later comparing or combining them during analysis and interpretation.^{14,25} The embedded design falls under convergent because it also collects qualitative and quantitative data simultaneously, but places more emphasis on one method, and uses the other method on a subset to the overall study. The two sequential designs aim to use one method to inform or explain the other.^{14,25} The explanatory sequential design collects and analyzes quantitative data during a first phase and uses qualitative methods in a second phase to explain the quantitative results.^{14,25} In contrast, the exploratory sequential design starts with a qualitative phase to explore a topic and informs a second quantitative phase.^{14,25}

Some epidemiologists already use mixed methods in epidemiology, particularly when developing surveys,^{26–28}

and other epidemiologists may incorporate aspects of mixed methods in their studies, but not formally or explicitly. For instance, epidemiologists may speak with members of the population when designing studies or include interpretations derived from observations during field work in the Discussion sections of manuscripts, yet they may not describe these qualitative details in the Method or Results sections, respectively. Some may argue that this is just what a good epidemiologist does to generate ideas or interpret data. Our rebuttal is: why must the qualitative aspects of what epidemiologists do be buried in their toolbox? We now turn to how epidemiologists can systematically apply mixed methods to the epidemiologic research process.

Applications of Mixed Methods to the Epidemiologic Research Process

The 2-by-2 table is at the core of epidemiology and mixed methods can help epidemiologists think through what belongs in that table (exposure and outcome) and what matters outside of the table when it comes to confounding, selection bias and attrition, measurement, mediation, and effect modification. We consider the first three of these concepts as causal identification (identifying potential causes and eliminating alternative explanations) and the latter two as causal explanation (explaining how and under what circumstances causes operate).²⁴ Figure 1 summarizes which mixed-methods study designs are best suited to strengthen each aspect of observational studies. The Table provides further details including mixed-methods examples for each epidemiologic concept. The best mixed-methods study design depends upon which aspect of the research question the epidemiologist chooses to enhance.

Association

When identifying potential causes of disease, qualitative methods allow epidemiologists to make observations of the population or other key stakeholders to generate new, grounded²⁹ hypotheses. An exploratory sequential design, including interviews, and observations in the qualitative phase, might aid epidemiologists to identify potential causes in the following ways: first, participants may describe a phenomenon not found in previous literature; and second, qualitative data focusing on the cultural context may identify upstream factors, such as family- or society-level determinants of disease, to be conceptualized as new potential causes. In the Table, we provide a hypothetical example of how interviewing school employees and observing children in schools helps an epidemiologist to identify pollution from a new food factory, and a food allergen in the snack the factory makes, as potential causes of high rates of asthma in a specific school district.

Confounding

Going to the population under study and gleaning on-the-ground perspectives can help epidemiologists understand

how to make the exposed and unexposed in their sample less confounded. Specifically, using an exploratory sequential design, participant observation and in-depth interviews can help epidemiologists identify respective community and individual level factors that may confound the main association. Qualitative data may also reveal social processes that connect variables to each other to help identify confounding. In the Table, we provide a hypothetical example to understand if social support, sex education, and body positivity can help determine if maternal history of menstrual pain is a confounder in the association between a diagnosis of fibroids and substance abuse.

Selection Bias into a Study

Qualitative methods can identify what factors to compare between the study population and study sample to assess selection bias that occurs during recruitment of participants into a study. Either exploratory sequential or convergent designs may be used. An example of the latter comes from Gallaway et al. who conducted a mixed-method case-control study of risky and protective factors of suicide in soldiers.³⁰ They used medical records, surveys, interviews, and focus groups and found similar demographic and military characteristics between soldiers who died by suicide versus accidental death. We see expanding upon their convergent design to further assess selection bias by interviewing soldiers about reasons for substance use, a major risk factor for suicide. The study team could compare the distribution of baseline factors by sub-groups based on the reasons for substance use, and then, the subgroup distributions to the overall sample. If the distributions are similar, this would suggest little to no selection bias based on substance use. If one subgroup was more similar to the overall distribution, this would point to selection bias and could help determine to which subgroup of soldiers the results may be most generalizable.

Attrition/Loss to Follow-up

Qualitative methods can help assess attrition, another potential source of selection bias, both during a study and after its completion. Epidemiologists can collaborate with ethnographers in convergent or embedded designs to understand the dynamics of a quantitative research study including the study rationale, design, recruitment, retention and role of study personnel, participants, and advocates. As we explain in hypothetical convergent and explanatory sequential examples in the Table 1, an ethnographer finds a conflict between the university and a local politician over gentrification that influences retention into a study. This information can help an epidemiologist determine if those lost to follow-up is a source of bias.

Measurement

Minimizing measurement error of key variables is essential in epidemiology.³¹ Epidemiologists might employ

TABLE. Combining Core Epidemiologic Concepts and Mixed Methods

Epidemiologic Concept	Qualitative Component	Quantitative Component	Mixed-Method Study Design	Examples
Association	Identify potential new causes of disease	Measure and test exposure and outcome in a quantitative phase	Exploratory Sequential	High rates of asthma are observed in students attending schools within the southeast region of a large US city, and the rates cannot be explained by existing risk factors. An epidemiologist's study begins with qualitative data collection using interviews with various key informants including city epidemiologists, clinicians, school principals, parents, and participant observation with students during the school day. There is a lot of mention about a new food factory that was recently built in that district and a viral social media video has many students consuming the new crunchy snack made by the factory. It seems that the added pollution and a possible food allergen in the snack may be triggering the high rates of asthma. The epidemiologist then tests this hypothesis in a quantitative phase in city schools.
Confounding	Identify confounders or connections between variables	Control for confounders in statistical model	Exploratory Sequential	An epidemiologist is interested in opioid use as pain management in those with fibroids, a very painful disease that is often undiagnosed. She designs a study beginning with in-depth interviews with women diagnosed with fibroids. When coding the interviews for themes, she generates a list of variables that may be related to either the exposure, outcome, or both. These variables, such as social support, previous sexual education, body positivity, maternal history of menstrual pain may be potential confounders or they may also help describe how confounders in the causal diagram are related to each other. The epidemiologist then builds a directed acyclic graph and decides which variables remain, and then tests it using quantitative methods.
Mediation and Effect modification	Identify possible mediators or effect modifiers	Examine association between exposure, mediator, modifier, and outcome	Exploratory Sequential	An epidemiologist is interested in knowing why an area of Miami has higher rates of cervical cancer. The exploratory sequential design begins with participant observation and in-depth interviews with Haitian women who predominantly live in that area, which reveals that twalet deba, a culturally mediated feminine hygiene practice, is widespread and many local shops sell products, both natural and synthetic, for twalet deba. ³⁴ In collaboration with community partners the epidemiologist begins to think that certain products used in twalet deba may increase the risk of cervical cancer. They then use quantitative methods to test whether any of the intravaginal agents used in twalet deba are effect modifiers for high risk HPV, a precursor for cervical cancer risk. ³³
	Gain an in-depth perspective of the mediators, or effect modifiers, causal partners	Examine association between exposure, mediator, modifier, and outcome	Explanatory Sequential	An epidemiologist is interested in the high rate of tuberculosis on indigenous population of Colombia. ⁴⁷ He uses quantitative methods to identify gaps in the tuberculosis care cascade and then uses interviews to explore why the gaps in the cascade are there.
	Identify possible mediators, or effect modifiers, causal partners of the main effect	Examine association between exposure, mediator, modifier, and outcome	Convergent	An epidemiologist is interested if a higher socioeconomic status (SES) and body mass index (BMI) are positively associated in a new context and want to explore socio-behavioral variables associated with BMI. ⁴⁸ She quantitatively measures SES, BMI, and other variables such as marital status, number of food market visits using structured questionnaires. She uses multivariate linear regression to confirm that SES is positively associated with BMI. Her collaborator collects qualitative data through participant observation and spends several weeks living and interacting with community members. In addition, he conducts in-depth interviews on food shopping, preparation, presentation, and nutrition perspectives. Families from higher SES tended to live closer to the market, which in turn seemed to lead to a higher number of market visits with an increased consumption of ultra-processed foods. In this community with a high prevalence of undernutrition during childhood, participants showed a preoccupation with hunger, rather than obesity. The qualitative component not only aligns with the quantitative finding that a higher SES is associated with higher BMI, but also provides details on how the exposure and outcome are linked in this particular setting.

Measurement	Identify constructs and ways to inquire about them to develop a survey	Use survey results to measure variables	Exploratory Sequential	An epidemiologist wants to measure depression in older adults in a future epidemiologic study. First, to better measure depression, she wants to understand the definition of depression in older adults in contrast to clinical definitions. ²⁷ So her collaborators conduct semistructured interviews and asks participants to describe "a person who is depressed." The qualitative results suggest that traditional measures of depression do not capture loneliness, but it was one of the most salient terms of the definition from older adults. For the subsequent quantitative phase, they develop a survey that measures depression that includes loneliness.
	Explore a variable to complement, expand, contrast how it is measured quantitatively	Measure a variable to complement, expand, contrast how it is measured quantitatively	Convergent	An epidemiologist designs a study to examine the association between discrimination and cardiovascular risk in Black men. He uses an existing questionnaire on discrimination to survey participants. At the same time he knows that the survey may not capture discrimination accurately so he also conducts focus groups with participants.
Selection Bias	Observe population for characteristics	Measure study sample characteristics	Exploratory Sequential	An epidemiologist is conducting a study of pregnancy associated breast cancer. Her co-investigator observes moms in parks and attends mom groups/activities for moms and babies in the same catchment area for the study. She notices at first there is a good representation of moms in terms of where they were born, but that after 2 weeks, most of the migrant women are no longer attending activities. The study team realizes many of the migrant moms have returned to work and this may affect their ability to keep breastfeeding. She makes sure to recruit moms from different types of work into the study to avoid selecting only moms with jobs that support breastfeeding. She decides to measure work-based policies for family leave and breastfeeding support at baseline to determine if she has selected women with different work-based lactation support. This approach helps identify covariates to measure to assess selection bias.
	Capture diverse perspectives on a variable/topic of study and use these perspectives to classify survey respondents from the QUANT component	Measure variables and compare their distribution by perspectives determined using QUAL methods	Convergent	An epidemiologist is interested in conducting a study of suicide among soldiers. He conducts a case-control study of suicide deaths using military clinical chart review. He also administers survey and focus groups interviews with soldiers and healthcare providers to provide context for associations between identified risk factors. ³⁰ To assess selection bias in the survey, the research team could also interview soldiers about substance abuse and identify codes (reasons for substance abuse) and sub codes (chronic pain, mental health, etc). The interviewees also complete the quantitative survey. Then, the epidemiologist classifies interviewees by sub codes and compares the distribution of substance use between these groups and the survey study sample. If the distribution differs between the quantitative sample and any of the qualitative subgroups, there may be selection bias based on what defines that particular subgroup.
Attrition/Lost to Follow-up	Understand the dynamics of the research study, including the study rationale, design, recruitment, retention and role of study personnel, participants, and advocates.	Examine association of exposure and outcome using a cohort design	Convergent/ Embedded	An epidemiologist conducts a cohort study looking at the association of socioeconomic status (SES) and early-onset breast cancer risk (<40 years) in Mexican migrants to the United States. She assesses SES using a survey at baseline and follow women for 10 years for breast cancer outcomes. Over those same 10 years an ethnographer follows the study, its participants, and the epidemiologist. In year 7, the ethnographer picks up on a political issue that may affect retention. The university and a prominent local politician are in conflict about gentrification and the politician is dissuading residents from engaging with the university. The epidemiologist and ethnographer hold a town hall to hear the community's concerns. They also incorporate zip code to account for loss to follow-up in the statistical analysis.
	Interview participants about the characteristics that differed between those lost to follow-up and who stayed on study	Compare baseline characteristics of those lost to follow with those with complete follow-up	Explanatory Sequential	Using the same example as above, an epidemiologist assesses socioeconomic status using a survey at baseline and follow women for 10 years for breast cancer outcomes. When she is creating Table 1 of study participants, she notices those lost to follow-up were more likely to live in a particular zip code. She decides to interview participants who completed follow-up living in that same zip code and come to learn that a local politician urged residents to stop participating in the University's studies. The interviews help the epidemiologist decide whether those lost to follow-up differ from the overall sample regarding socioeconomic status and how to account for this in the statistical analysis.

qualitative methods first to decide how or what variables to measure quantitatively or they may use both quantitative and qualitative methods concurrently to measure variables.¹⁶ Qualitative methods can assist epidemiologists to identify language and colloquialisms to measure a variable or ways to phrase questions about potentially sensitive topics. When a previously validated instrument needs to be translated and validated in a different context, qualitative methods ground the necessary changes in an emic perspective.¹⁶ For example, when one of us (LCH) was working with an interdisciplinary team to reconstruct the Native American diet consumed in the 1940s in New Mexico, some collaborators did not see it necessary to ask about dairy intake because of the documented high prevalence of lactose intolerance in Native Americans. However, when we interviewed Native American elders about dairy intake as children, they recalled their family members eating dairy and one participant remembered his family making cheese. Sequential exploratory designs are most suited for when epidemiologists want to use qualitative methods to design survey instruments. For example, Barg et al.²⁷ explored the meaning of depression in older adults and came to learn that loneliness was a major part of that definition before creating a survey that incorporated loneliness into the depression measure (Table). If epidemiologists want to measure the same variable using both quantitative and qualitative methods, a convergent design is appropriate (Table).

Mediation and Effect Modification

Mixed methods can assist epidemiologists to understand how an association works (mediation) from the perspective of the population under study, or to capture context (group-level, cultural, or social factors) for identifying effect modifiers. Causation cannot be completely isolated from context, as one factor might be a causal factor of an outcome in one environment but not in another one, depending on the distribution of causal partners in each setting.³² While quantitative methods can describe the quantitative distribution of causal partners, qualitative methods can inform how and why the relationship between exposure and outcome differs between contexts. Either sequential or convergent designs may be useful in explaining causal mechanisms (Table). An example of using mixed methods to identify an effect modifier comes from Erin Kobetz's work which found that *twalet deba*, a culturally mediated feminine hygiene practice among many Haitian women, may explain high rates of cervical cancer in Little Haiti in Miami.^{33,34} Although this research was not a single mixed-methods study, the authors used qualitative results to inform the quantitative test of whether intravaginal agents increased susceptibility to cervical cancer.

We have discussed applications of mixed methods to enhance hypothetical and current epidemiologic studies by aligning mixed-methods study designs with epidemiologic concepts. Although it is common to use the term "mixed methods" when referring to studies using at least one quantitative and one qualitative method, the purpose of mixed

methods is to integrate multiple methods during interpretation. There are many examples of mixed-methods studies that use qualitative data to develop an epidemiologic survey²⁶ and collect qualitative data to understand perspectives of disease outcomes.³⁵ There are fewer examples of epidemiologic studies that also integrate results during the analysis phase.^{36,37} We now describe a case study that exemplifies mixed-methods integration in observational epidemiology.

Case Study of an Epidemiologic, Observational Study using Mixed-methods

To better understand what early life factors explain rising breast cancer incidence rates among migrants that move from low to high incident countries, we conducted a mixed-methods migrant study on puberty.^{38–40} Earlier age at puberty is associated with increased breast cancer risk,⁴¹ so we compared pubertal timing within the context of migration.⁴⁰ To align with literature on puberty and breast cancer, we measured puberty following biomedical definitions and used established epidemiologic methods (validated questionnaires and hormonal biomarkers).^{42,43} At the same time, given the inclusion of different cultural groups in our sample (White British girls and British–Bangladeshi migrants in London, UK, and Bangladeshi girls in Sylhet, Bangladesh) we used qualitative methods to understand the context in which girls were growing up. From the literature we knew that body mass index (BMI) was a potential mediator, but we were interested in identifying other mediators from an emic perspective. Therefore, our research question necessitated mixed methods.

We followed a convergent design to assess biocultural constructs related to both migration (exposure, X) and puberty timing (outcome, Y). Figure 2 illustrates the causal diagram and uses color to indicate the quantitative and qualitative methods to measure each variable. Quantitative data collection involved measuring puberty (Y) through a hormonal biomarker and the Pubertal Development Scale.⁴² A structured questionnaire assessed aspects of migration (X), such as preference for clothes and food.⁴⁰ To calculate BMI (mediator, M) we took anthropometric measurements. The qualitative data collection occurred during afterschool clubs, and included participant observation and focus groups to gather girls' perspectives of social expressions of puberty (Y), such as choice of clothes and wearing the hijab, and food preferences, specifically eating rice and curry, which was both a marker of migration (X) and related to BMI (M).⁴⁰ We collected qualitative and quantitative data in parallel and placed equal emphasis on the qualitative and the quantitative components.

Quantitative analysis included survival models to compare age at puberty among White British, migrant and Bangladeshi girls, as well as mediation by BMI.³⁸ Analysis of qualitative data included open coding and grounded theory to analyze field notes and focus group discussions related to hijab and food. We used joint display,⁴⁰ an approach used to present qualitative and quantitative results simultaneously.⁴⁴ In a table, the first column displayed the quantitative results for

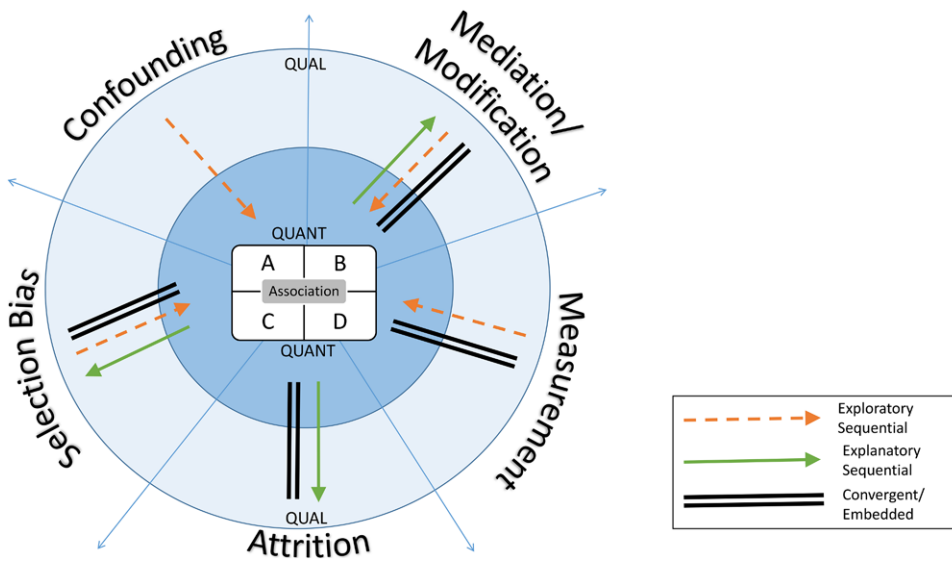


FIGURE 1. Mixed-method study designs aligned with epidemiologic concepts. Mixed methods can help epidemiologists incorporate the emic view into many aspects of research while building causal models. The 2×2 table is the foundation of epidemiology and quantitative at the core. Qualitative methods can be incorporated either before, during, or after estimating the association between exposure and outcomes. The order in which the quantitative and qualitative methods are used depend on what aspect of the research question needs to be strengthened.

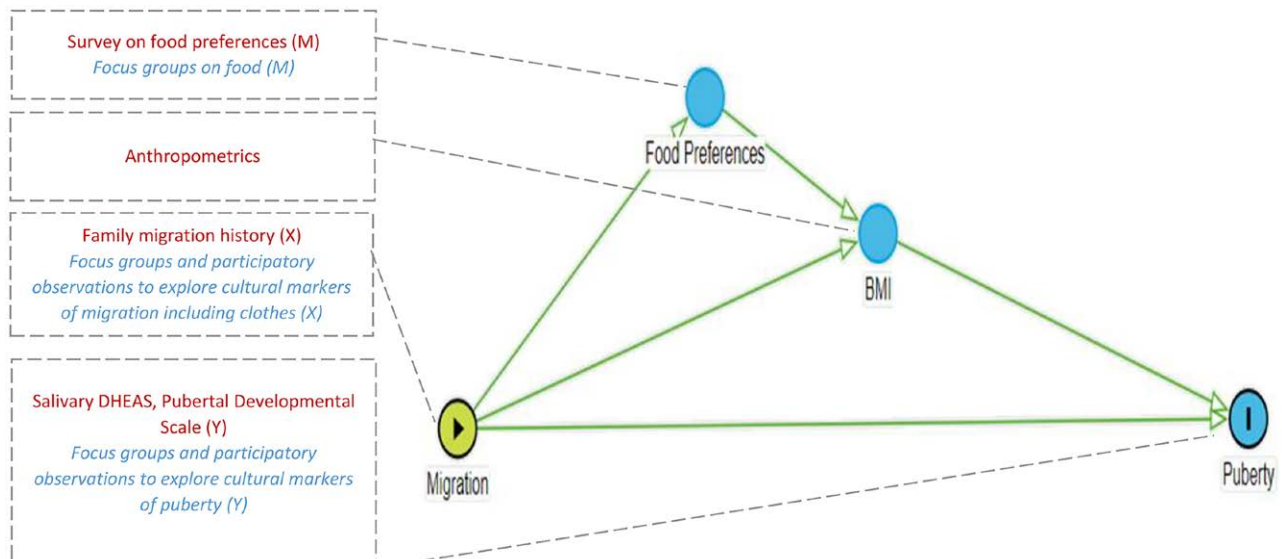


FIGURE 2. Hypothesized DAG with associated qualitative (italics) and quantitative (Roman) methods in mixed-methods Convergent Study, Adolescence among Bangladeshi and British Youth.

each study variable through bar charts, the second and third columns presented corresponding quotes from Bangladeshi and migrant participants, respectively. The joint display highlighted where the biologic and cultural definitions of each variable converged or diverged. For example, girls reported eating rice and curry for dinner in 24-hour food recalls, but in the same day said to their friends, “I don’t eat rice,” which was a way to express rejection of Bangladeshi culture.

The quantitative results confirmed that migrant girls experienced puberty earlier than nonmigrant girls.³⁸ BMI partially explained the association between migrant group (X) and puberty timing (Y). Qualitative data suggested 1st generation migrant girls, the group with earliest pubertal age,

experienced discrimination and stress.⁴⁰ Our use of mixed methods allowed for the integration of data in a way we had not initially planned. Early on during field work, we noticed that some girls did not wear hijab every day. We were perplexed as we thought this was a rather fixed cultural practice. However, girls explained, “I’m only practicing, I’m not yet dedicated to the scarf.” We revised our survey to ask girls if they wore the scarf occasionally or every day and used this dichotomous variable as an additional pubertal outcome in survival models. We compared the median age at pubertal onset between our biologic and cultural definitions and found that “practicing” aligned with the hormonal rise in androgens around age 5 (adrenarche) and “being dedicated” aligned with

the age at menarche in migrant Bangladeshi girls.⁴⁰ This integrated analysis illustrated the relationship between social and biologic markers of puberty, which was a contribution beyond previous studies that investigated social and biologic factors of puberty separately.

We have illustrated the alignment of mixed-methods design with epidemiologic concepts through examples and a case study. Now, we will turn to an application that cuts across the epidemiologic concepts, which entails using mixed methods to define causal structures.

MIXED METHODS TO DEFINE CAUSAL STRUCTURES

Defining the underlying causal structure of a phenomenon in epidemiology entails identifying causes of health outcomes and describing how and for whom the associations between causes (exposures, X) and health outcomes (Y) work.³¹ Causal diagrams including but not limited to directed acyclic graphs (DAGs) are one way of illustrating the underlying causal structure. However, epidemiologists predominantly build DAGs using their etic perspective, external to the population under study. Combining the etic with the emic—insider perspective of the context within which the phenomenon occurs—provides a new approach to building DAGs.

A challenge when constructing a useful and meaningful DAG is understanding when two nodes are related, the direction of the arrow, and whether a covariate might be a confounder, mediator, collider, or irrelevant variable. Often there is a lack of theory and sufficient empirical data to be certain of these structures. Furthermore a DAG cannot provide insight into what variables may be missing or whether a variable is conceptualized appropriately.⁹ Qualitative data can provide additional empirical data defining the underlying structure of causal relationships. During qualitative data analysis, mapping options in qualitative coding software, such as NVivo,⁴⁵ help to identify important nodes and the meaningful connections between them, in a similar way as building a DAG. In NVivo, nodes are qualitative parent and child codes that researchers generate either deductively—the researcher searches for text relating to a preconceived code—or inductively—the code emerges from textual data. Qualitative methods offer a DAG the meaning of variables and connections between them from an emic perspective. Figure 3 shows a sequential exploratory study that collects qualitative data from women with early-onset breast cancer to build a causal diagram to test with quantitative methods. Qualitative analysis identifies parent codes (Air pollution, Stress, Marital Status) as possible causes of cancer. In telling their story of getting early-onset breast cancer, women said “I found a lump while on honeymoon” or “I thought it was related to breastfeeding” and such qualitative data yield two child codes, Parity and Breastfeeding, under Marital Status (Figure 3A). These five codes become variables in a DAG (Figure 3B) and qualitative data, as well as evidence from previous studies, inform the connections between them.

The epidemiologist can test the idea that breastfeeding is positively associated with early-onset cancer, an idea that they may not have had before interviewing women since breastfeeding is negatively associated with postmenopausal breast cancer.

We recognize that triangulation in epidemiology often implies comparing results across more than one study. Returning to the Adolescence among Bangladeshi and British Youth case study, Figure 4 illustrates how using mixed methods within a single study can define underlying causal structures for future studies. Qualitative information on discrimination and stress, such as “I’m not a Freshie” and “I’m proud of my religion but not my culture,” helped inform questions as to why puberty was particularly early in first-generation migrants. BMI and stress are established risk factors for early puberty, but seldom analyzed as causal partners, thus mixed methods led us to a new DAG that includes a hormonal mechanism for the interaction between stress and BMI (Figure 4).

LIMITATIONS OF INTEGRATING MIXED-METHODS AND EPIDEMIOLOGY

We recognize limitations of applying mixed methods in epidemiology at the present time. With no formal training in mixed methods, current epidemiology teams may lack expertise and will need new collaborations with qualitative researchers. Yet lack of training does not preclude epidemiologists from designing mixed-methods studies. We envision epidemiologists who can design their own mixed methods epidemiologic studies and then collaborate with experienced qualitative researchers to conduct the research. Mixed-methods studies may require more time and resources than studies only using quantitative methods and securing funding for epidemiology studies using mixed-methods may be difficult. However, the Office of Behavioral and Social Sciences Research at the National Institutes of Health commissioned the “Best Practices for Mixed Methods Research in the Health Sciences” to assist investigators, reviewers and NIH leadership.⁴⁶ Last, despite carefully planned designs, there may be situations where data cannot be easily integrated or provide opposing conclusions. We have had this experience but found that divergent results lead to new hypotheses.

CONCLUSION

Krieger stated that an “intellectual and empirical challenge is to integrate biomedical, lifestyle and social risk factors to afford a richer understanding of the causal processes at play and hence better inform efforts to improve population health and reduce health inequities.”⁴⁷ We argue that mixed methods allows for the integration of bio-socio-cultural factors in epidemiologic studies. We align mixed-methods study designs with epidemiologic concepts so that epidemiologists can enhance observational studies. We describe how mixed methods can define the underlying causal structure of phenomenon. Our how to guide overcomes a major critique of efforts to improve causal inference that epidemiology

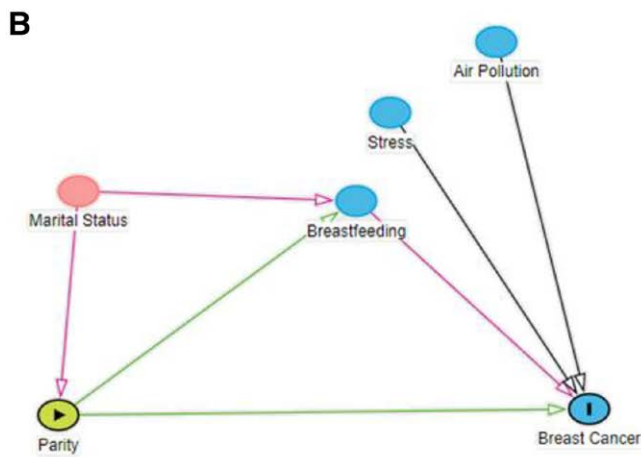
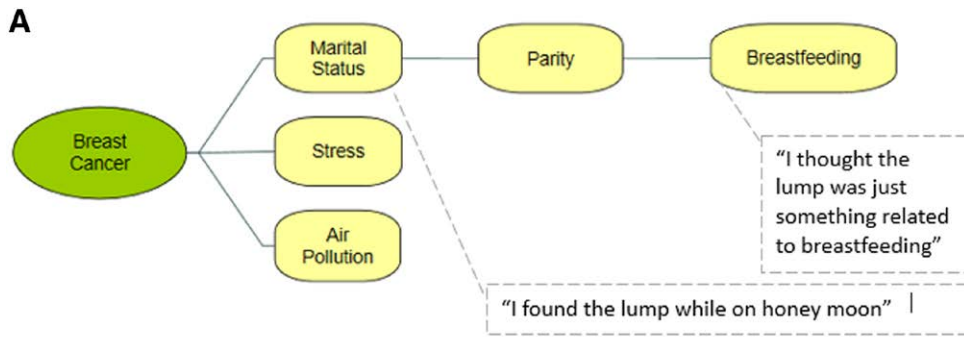


FIGURE 3. Using qualitative analysis in NVivo (A) to inform DAGs (B) within a sequential exploratory mixed-method study design.

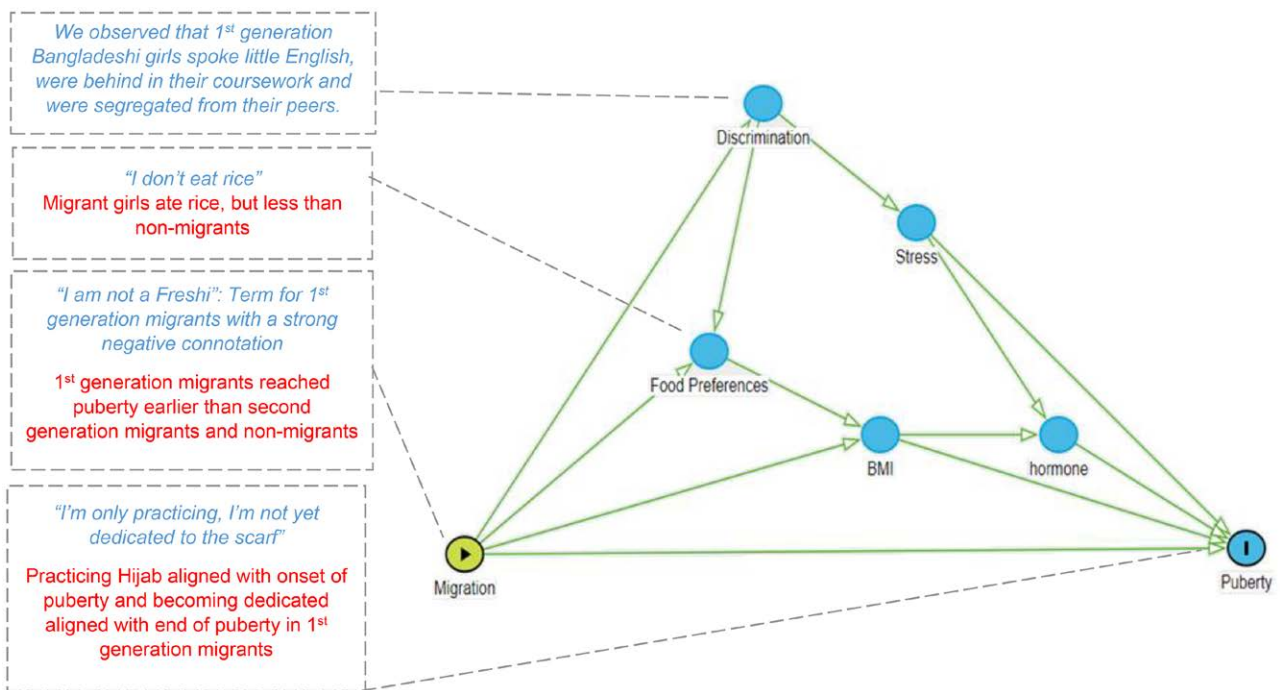


FIGURE 4. Updated DAG informed by qualitative (*italics*) and quantitative (Roman) results from a mixed-methods convergent study, Adolescence among Bangladeshi and British Youth.

textbooks currently do not include.⁴⁷ Mixed methods is a systematic approach to determining what goes into our causal structures. Previously hidden in the causal inference toolbox, we have described how to systematically incorporate the perspective and context of the population under study and how to integrate the social and biological factors of health and diseases within single epidemiologic studies.

ACKNOWLEDGMENTS

We would like to thank Dr. Sharon Schwartz for her helpful feedback on earlier drafts of this article and Hanfei Qi for developing related web content that helped reorganize the current article.

REFERENCES

- Ben-Shlomo Y, Kuh D. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *Int J Epidemiol*. 2002;31:285–293.
- Tsai AC, Mendenhall E, Trostle JA, Kawachi I. Co-occurring epidemics, syndemics, and population health. *Lancet (London, England)*. 2017;389:978–982.
- Singer M, Bulled N, Ostrach B, Mendenhall E. Syndemics and the biosocial conception of health. *Lancet (London, England)*. 2017;389:941–950.
- Rothman KJ, Greenland S. Causation and causal inference in epidemiology. *Am J Public Health*. 2005;95:S144–S150.
- Hernán MA, Robins JMJ. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC. 2020. Available at: <https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/>. Accessed August 25, 2021.
- Morabia A. Has epidemiology become infatuated with methods? A historical perspective on the place of methods during the classical (1945–1965) phase of epidemiology. *Annu Rev Public Health*. 2015;36:69–88.
- DiGiacomo SM. Can there be a “cultural epidemiology?” *Med Anthropol Q*. 1999;13:436–457.
- Bach M, Jordan S, Hartung S, Santos-Hövenner C, Wright MT. Participatory epidemiology: the contribution of participatory research to epidemiology. *Emerg Themes Epidemiol*. 2017;14:2.
- Krieger N, Smith GD. The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology. *Int J Epidemiol*. 2016;45:1787–1808.
- Susser M. Should the epidemiologist be a social scientist or a molecular biologist? *Int J Epidemiol*. 1999;28:S1019–S1022.
- Schwartz S, Gatto NM, Campbell UB. Causal identification: a charge of epidemiology in danger of marginalization. *Ann Epidemiol*. 2016;26:669–673.
- Vandenbroucke JP, Broadbent A, Pearce N. Causality and causal inference in epidemiology: the need for a pluralistic approach. *Int J Epidemiol*. 2016;45:1776–1786.
- Johnson RB, Russo F, Schoonenboom J. Causation in mixed methods research: the meeting of philosophy, science, and practice. *J Mix Methods Res*. 2019;13:143–162.
- Creswell JW, Plano Clark VL. *Designing and Conducting Mixed Methods Research*. Third Edit. Los Angeles: SAGE Publications; 2018.
- Sale JEM, Lohfeld LH, Brazil K. *Revisiting the Quantitative-Qualitative Debate: Implications for Mixed-Methods Research*. Vol 36. Qual Quant; 2002.
- Curry LA, Nembhard IM, Bradley EH. Qualitative and mixed methods provide unique contributions to outcomes research. *Circulation*. 2009;119:1442–1452.
- Bogner HR, Dahlberg B, De Vries HF, Cahill E, Barg FK. Older patients’ views on the relationship between depression and heart disease. *Fam Med*. 2008;40:652–657.
- Johnson RB, Onwuegbuzie AJ. Mixed methods research: a research paradigm whose time has come. *Educ Res*. 2004;33:14–26.
- Palinkas LA, Mendon SJ, Hamilton AB. Innovations in mixed methods evaluations. *Annu Rev Public Health*. 2019;40:423–442.
- Bazeley P. Issues in mixing qualitative and quantitative approaches to research. In: Buber R, Gadner J, Richards L, eds. *Applying qualitative methods to marketing management research*. Palgrave Macmillan; 2004:141–156.
- Russo F, Williamson J. Interpreting causality in the health sciences. *International Studies in the Philosophy of Science*. <http://dx.doi.org/10.1080/02698590701498084>. 2007;21:157–170.
- Reiss J. Causation in the sciences: an inferentialist account. *Stud Hist Philos Sci Part C Stud Hist Philos Biol Biomed Sci*. 2012;43:769–777.
- Bannister-Tyrrell M, Meiqari L. Qualitative research in epidemiology: theoretical and methodological perspectives. *Ann Epidemiol*. 2020;49:27–35.
- Shadish WR, Cook TD, Campbell DT. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin; 2002.
- Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs—principles and practices. *Health Serv Res*. 2013;48(6pt2):2134–2156.
- Shariff-Marco S, Gee GC, Breen N, et al. A mixed-methods approach to developing a self-reported racial/ethnic discrimination measure for use in multiethnic health surveys. *Ethn Dis*. 2009;19:447–453.
- Barg FK, Huss-Ashmore R, Wittink MN, Murray GF, Bogner HR, Gallo JJ. A mixed-methods approach to understanding loneliness and depression in older adults NIH public access. *J Gerontol B Psychol Sci Soc Sci*. 2006;61:S329–S339.
- Quinn M, Cummings C, Stinson J. 457A mixed methods approach to understanding Adverse Childhood Experiences (ACEs) in Munsieville, South Africa. *Int J Epidemiol*. 2021;50(Supplement_1):dyab168.551.
- Glaser BG, Strauss AL. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Available at: <https://books.google.com/books?id=C5QiwAEACAAJ&dq=editions:5HS7DU010egC&hl=en&sa=X&ved=0ahUKEwjXxPC6svjhAhUDneAKHRatCwsQ6AEIUDAG>. Accessed April 30, 2019.
- Galloway MS, Lagana-Riordan C, Dabbs CR, et al. A mixed methods epidemiological investigation of preventable deaths among U.S. Army soldiers assigned to a rehabilitative warrior transition unit. *Work*. 2015;50:21–36.
- Savitz DA, Wellenius GA. *Interpreting Epidemiologic Evidence: Connecting Research to Applications—David A. Savitz, Gregory A. Wellenius—Google Books*. 2nd edn. Oxford University Press; 2016.
- Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd edn. Lippincott Williams & Wilkins; 2008.
- Seay JS, Mandigo M, Kish J, Menard J, Marsh S, Kobetz E. Intravaginal practices are associated with greater odds of high-risk HPV infection in Haitian women. *Ethn Health*. 2017;22:257–265.
- Menard J, Kobetz E, Diem J, Lifleur M, Blanco J, Barton B. The sociocultural context of gynecological health among Haitian immigrant women in Florida: applying ethnographic methods to public health inquiry. *Ethn Health*. 2010;15:253–267.
- Galson SW, Staton CA, Karia F, et al. Epidemiology of hypertension in Northern Tanzania: a community-based mixed-methods study. *BMJ Open*. 2017;7:e01882918829.
- Zhang W, Creswell J. The use of mixing procedure of mixed methods in health services research. *Med Care*. 2013;51:e51–e57.
- Cai S, Wang N, Xu L, et al. Impacts of antibiotic residues in the environment on bacterial resistance and human health in Eastern China: an interdisciplinary mixed-methods study protocol. *Int J Environ Res Public Health*. 2022;19:81458145.
- Houghton LC, Cooper GD, Bentley GR, et al. A migrant study of pubertal timing and tempo in British-Bangladeshi girls at varying risk for breast cancer. *Breast Cancer Res*. 2014;16:469.
- Houghton LC, Cooper GD, Booth M, et al. Childhood environment influences adrenarheal timing among first-generation Bangladeshi migrant girls to the UK. *PLoS One*. 2014;9:e109200.
- Houghton LC, Troisi R, Sommer M, et al. I’m not a freshie: Culture shock, puberty and growing up as British-Bangladeshi girls. *Soc Sci Med*. 2020;258:113058.
- Bodicoat DH, Schoemaker MJ, Jones ME, et al. Timing of pubertal stages and breast cancer risk: the Breakthrough Generations Study. *Breast Cancer Res*. 2014;16:R18.

42. Brooks-Gunn J, Warren MP, Rosso J, Gargiulo J. Validity of self-report measures of girls' pubertal status. *Child Dev.* 1987;58:829–841.
43. Petersen AC, Crockett L, Richards M, Boxer A. A self-report measure of pubertal status: Reliability, validity, and initial norms. *J Youth Adolesc.* 1988;17:117–133.
44. Guetterman TC, Fetters MD, Creswell JW. Integrating quantitative and qualitative results in health science mixed methods research through joint displays. *Ann Fam Med.* 2015;13:554–561.
45. Qualitative Data Analysis Software | NVivo. <https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home>. Accessed February 22, 2022.
46. Meissner H, Creswell J, Klassen AC, Plano V, Smith KC. Best Practices for Mixed Methods Research in the Health Sciences. Office of Behavioral and Social Sciences Research. National Institutes of Health. Available at: https://obsr.od.nih.gov/sites/obsr/files/Best_Practices_for_Mixed_Methods_Research.pdf. 2011. Accessed 20 July 2021.
47. Márquez IAR, Hoyos KYT, Pereda M del PT, Salazar BLG, Pérez F, Pasaje JEP. 355 Tuberculosis care cascade in an indigenous population of a Colombian region. *Int J Epidemiol.*
48. Nagata JM, Vallengia CR, Barg FK, Bream KDWW. Body mass index, socio-economic status and socio-behavioral practices among Tz'utujil Maya women. *Econ Hum Biol.* 2009;7:96–106.
49. Krieger N. *Epidemiology and the People's Health: Theory and Context*. Oxford University Press; 2011.